A Novel Design and Implementation of DoS-Resistant Authentication and Seamless Handoff Scheme for Enterprise WLANs

---

A thesis

submitted in partial fulfillment

of the requirements for the Degree

of Master of Science in Computer Science

in the University of Canterbury

by Isaac (Chien-Wei) Lee

University of Canterbury

Department of Computer Science

and Software Engineering

2010

Abstract

With the advance of wireless access technologies, the IEEE 802.11 wireless local area network (WLAN) has gained significant increase in popularity and deployment due to the substantially improved transmission rate and decreased deployment costs. However, this same widespread deployment makes WLANs an attractive target for network attacks. Several vulnerabilities have been identified and reported regarding the security of the current 802.11 standards. To address those security weaknesses, IEEE standard committees proposed the 802.11i amendment to enhance WLAN security. The 802.11i standard has demonstrated the capability of providing satisfactory mutual authentication, better data confidentiality, and key management support, however, the design of 802.11i does not consider network availability. Therefore, it has been suggested that 802.11i is highly susceptible to malicious denial-of-service (DoS) attacks, which exploit the vulnerability of unprotected management frames.

This research first investigates common DoS vulnerabilities in a Robust Security Network (RSN), which is defined in the 802.11i standard, and presents an empirical analysis of such attacks – in particular, flooding-based DoS attacks. To address those DoS issues, this thesis proposes a novel design and implementation of a lightweight stateless authentication scheme that enables wireless access points (APs) to establish a trust relationship with an associating client and derive validating keys that can be used to mutually authenticate subsequent layer-2 (link layer) management frames.

The quality of service provisioning for real-time services over a WLAN requires the total latency of handoff between APs to be small in order to achieve seamless roaming. Thus, this thesis further extends the proposed link-layer authentication into a secure fast handoff solution that addresses DoS vulnerabilities as well as improving the existing 802.11i handoff performance. A location management scheme is also proposed to minimise the number of channels required to scan by the roaming client in order to reduce the scanning delay, which could normally take up 90% of the total handoff latency.

In order to acquire practical data to evaluate the proposed schemes, a prototype network has been implemented as an experimental testbed using open source tools and drivers. This testbed allows practical data to be collected and analysed. The result successfully demonstrated that not only the proposed authentication scheme eradicates most of the DoS vulnerabilities, but also substantially improved the handoff performance to a level suitable for supporting real-time services.

Table of Contents

List of Figures

List of Tables

Previously Published Material

The following papers are the author's previous works that have been published or presented, and contain material that this thesis is based on.

- *Lee, I. and Hunt, R. (2008) 'A novel design of a VoIP firewall proxy to mitigate SIP-based flooding attacks', Int. J. Internet Protocol Technology, Vol. 3, No. 2, pp.128–135.*

- *Xianglin Deng and Chien-wei Lee (2008) 'Security of VoIP: SIP Flooding and its Mitigation', NZCSRSC'08 proceedings. http://nzcsrsc08.canterbury.ac.nz/site/proceedings/Individual_Papers/pg204_Security_of_VoIP_-_SIP_Flooding_and_its_Mitigation.pdf*

Acknowledgements

This thesis is the result of a challenging yet fruitful journey, upon which many people have given support and contribution. Without them, the successful completion of this work would not have been possible.

First and foremost, I would like to thank my supervisor, Associate Professor Ray Hunt, for his advice and guidance throughout my research. He gave me the freedom to pursue my research ideas and supported me with encouragement and valuable feedback.

I would also like to thank the University of Canterbury for the opportunity to work at the Internet Security Lab, where I have learnt a lot from many of my colleagues. The resource and equipment supply from the Computer Science Department have also been a great help towards my research.

I am most grateful to our Heavenly Father for his amazing guidance, for being my source of strength, and giving me the determination to overcome in difficult and challenging times. I also wish to thank my friends and church members for their continuing support and prayers.

Special thanks go to my partner Jo-Ying Huang, for her loving support, understanding, and always being there for me.

Last but not least, my deepest gratitude goes to my family, especially my parents Alice and Barry, for their unconditional support and believing in me. It is to them that I dedicate this work.

Chapter 1

Introduction

1.1 Motivations

Wireless access technologies have been evolving in recent years. In particular, the prevalence of IEEE 802.11 [1] wireless local area networks (WLANs) has increased dramatically. Companies and enterprises are now deploying WLANs as corporate network extensions to enable workforce mobility and increase productivity. With such ubiquitous network connectivity, mobile users can access network resources and exchange information anywhere within the coverage.

Although WLANs have become a part of people's everyday life, the inherent security weaknesses and protocol flaws are still a major issue that needs to be addressed. Due to intrinsic characteristics of WLANs, attacks can be performed in a wireless medium remotely and over the air. Hence, unlike traditional wired network attacks, a WLAN attack does not need to gain physical access to the wires that transmits data; it only requires the attacker's device to be within reach of the wireless targets. Further, an attack may not only aim at a single wireless device, but can be repeated towards all devices within reach.

The IEEE 802.11 standard initially addressed security issues with the Wired Equivalent Privacy (WEP) protocol, which proved to have several security vulnerabilities [2, 3]. The IEEE802.11 Task Group I (TGi) later in 2004 developed an amendment to the standard, IEEE 802.11i [4], providing medium access control (MAC) security enhancements, including the robust security network (RSN). Not only has

IEEE 802.11i fixed most of the WEP vulnerabilities, but it also demonstrated the satisfactory mutual authentication, data confidentiality and key management in a RSN.

Any activity that prevents authorised users from performing appropriate functions may be considered a denial of service (DoS) attack. Overwhelming a victim's resources by flooding it with malicious traffic is the most basic and probably the most difficult type of DoS attack to defend, particularly in WLANs. In order to create maximum damage, most flooding-based DoS attacks not only try to cause damage by consuming the network bandwidth, but endeavour to consume a substantial amount of computation resources. Flooding-based DoS attacks can particularly make drastic degradation upon network availability if the victim device under attack maintains state information about the attempted connections. Thus, for example, a wireless access point (AP) can rapidly become exhausted and fail to continue providing services.

Unfortunately, IEEE 802.11i only concentrates on securing the data frames used to transport higher layer protocol data, leaving the management frames used for connection administration unprotected. As a result, an 802.11i secured WLAN remains highly susceptible to malicious DoS attacks that exploit the unprotected management frames. For instance, DoS can be achieved by flooding spoofed management frames, which are unprotected and unauthenticated, to deauthenticate or disassociate a client (or station) from its associated AP. Such attacks exhaust network resources and prevent legitimate users accessing the network. Even with the new IEEE802.11w management frame protection techniques in place, flooding a large volume of management frames against an AP in a WLAN can still lead to a DoS condition in the network. The lack of availability protection and the weak confidentiality caused by unprotected management frames, coupled with the intrinsic ability to easily inject or flood malicious frames into the medium, threatens the WLAN security. This is specifically the area of protection to which this research is directed.

In an enterprise environment, roaming between APs has become one of the basic elements of a WLAN’s operation. With the emerging trend of running real-time multimedia applications, such as voice over IP (VoIP), over a WLAN, seamless mobility has also become one of the requirements in order to achieve satisfactory

quality of service (QoS). Unfortunately, the original WLAN design only provides primitive mobility support which is far below the expectation for seamless mobility.

The research presented in this thesis first proposes a novel authentication scheme that addresses the DoS vulnerabilities by introducing an efficient link-layer frame authentication mechanism, based on the technique of client puzzles. The scheme allows the client and the AP to mutually authenticate the management frames and other vulnerable link-layer protocol control frames, such as the EAP messages. Hence, most of the spoofing based DoS and flooding attacks can be mitigated with this proposed scheme.

Improvements to the existing 802.11i based handoffs are also a major focus of this research. A fast and seamless roaming solution is proposed to improve the intra-domain handoff performance by reducing both the scanning delay associated with the discovery phase searching for candidate APs and the delay of the re-authentication process with the new AP. The combination of the proposed solutions introduced in the thesis will provide an additional DoS resistance to the RSN as well as allowing secure and seamless handoffs with the performance capable for meeting the QoS requirements for real-time multimedia services in an enterprise WLAN environment.

1.2 Research Objectives and Outcomes

As DoS attacks become more and more common in WLANs, there have been an increasing number of publications on those issues and solutions to address DoS attacks [5-9]. Those publications raise the awareness of DoS attacks and provide helpful practices to enhance security. However, there are not many studies that actually provide effective solutions to address different types of flooding-based DoS attacks. Further, there appear to be no experimental studies of the effects of these DoS attacks on network availability, handoff performance, and user experience of time-sensitive applications.

The purpose of this research is to provide comprehensive studies of flooding-based DoS attacks in a RSN, and devise an authentication framework, along with an experimental testbed to study and quantify the performance of the solution, to mitigate DoS attacks while supporting secure seamless roaming. The goal is achieved through the following objectives:

1. *The development of a testbed framework for the analysis of DoS vulnerabilities in a RSN.* The purpose of this objective is to realise a RSN testbed network in which DoS attacks can be launched and the effects can be empirically studied. The testbed should be fully functional, and the implementation will be based on open source tools. The open source nature of the implementation enables further investigation and experimentation by other researches. This testbed provides a basis to perform vulnerability analysis of existing WLAN security, as well as a platform from which to design, implement, and evaluate new authentication and handoff solutions.

2. *The design and implementation of a link-layer authentication scheme to address possible DoS attacks.* Based on the outcomes of the vulnerabilities analysis, mitigation techniques are investigated. Once an appropriate technique to mitigate DoS vulnerabilities has been selected, the research pursues the development of a frame layer authentication scheme adopting the selected technique to support the existing RSN protocols in improving DoS resistance. A novel authentication scheme, Access Point Nonce (APN) authentication, is proposed that extends the existing RSN to form a much more DoS-resistant security framework. The APN authentication starts with a stateless authentication exchange based on client puzzles to bind a legitimate identity to a special security key material called identity token, which can be used to authenticate subsequent frames from the other party that participated in the APN authentication exchange. This scheme eliminates the most prominent DoS attacks based on injecting or flooding spoofed frames.

3. *The design and implementation of an intra-domain handoff model that extends the frame protection from the APN authentication to handoff sessions as well as*

enhancing handoff performance to support real-time services, such as VoIP. The integration of roaming support and the APN authentication scheme requires the investigation of different handoff models. Existing handoff solutions are studied and analysed for their suitability in supporting the frame layer authentication scheme. Where necessary, existing models are modified or enhanced to provide such support. The result is a proposed novel handoff scheme called Fast AP Transition Protocol (FATP). This scheme ensures that before the station (STA) moves to the new AP some pre-computed security keys are already shared between the new AP and the STA so that the re-authentication delay could be reduced. This is achieved with the concept of trust transfer which delegates authentication authority between APs. As the FATP scheme supports the re-generation of the security parameters used by the APN authentication, subsequent link-layer frames in the handoff sessions can still be authenticated using the refreshed keying material. With the FATP scheme, secure and seamless handoffs can be achieved to meet the QoS requirements of real-time multimedia services without compromising the security level of the IEEE 802.11i standard.

4. *The design and implementation of a selective channel scan mechanism based on a location management scheme to reduce scanning delay.* The long delays associated with channel scanning and probing in the discovery phase of a handoff has always been a major obstacle to seamless handoffs. In this research, some investigations have been done to improve the scanning performance. A scanning scheme called Location Management based Selective Scanning (LM-SS) is proposed that dynamically constructs and maintains an AP topology database, which will assist a roaming client to determine what channels are required to scan. A cross-layer probing technique is used in this scheme to also eliminate the waiting time for probe responses. With the LM-SS, the AP discovery process will not cause significant interruption to the ongoing data communications so that the QoS requirements can still be met during handoffs.

5. *The performance evaluation of the proposed APN authentication and the LM-SS + FATP handoff schemes for fast roaming.* The proposed solutions (2-4) are implemented in the testbed and experimental measurements are obtained to quantitatively evaluate the effectiveness and performance of those solutions.

1.3 Thesis Outline

This research aims to provide solutions for enterprise WLANs to address existing DoS vulnerabilities as well as enabling users with secure and seamless roaming capabilities for real-time multimedia applications. The structure of the thesis is organised as follows:

Chapter 2 first gives an overview of IEEE 802.11 WLAN technologies, architecture, and related security standards. Chapter 3 discusses the DoS vulnerabilities in the existing RSN, and presents some of the common link-layer DoS attacks in WLANs. Based on this vulnerability analysis, Chapter 4 further experimentally investigates the impact of those DoS attacks in a testbed. The implementation of the testbed and some experimental results showing the degradation of network performance and availability due to the DoS attacks are presented and discussed. Some mitigation requirements are also identified. In Chapter 5, some DoS mitigation techniques are discussed and the proposed APN authentication scheme is introduced. To deliver seamless roaming support in a RSN, Chapter 6 first surveys several existing handoff schemes, and provides a detailed analysis on their security and performance. Based on the result, the FATP handoff scheme is proposed. Detailed description of the scheme will be presented and the security analysis of the scheme will also be discussed. Chapter 7 will focus on the link-layer scanning phase of the handoff process and proposes a location management-based scanning scheme to shorten the latencies associated with the scanning activities. In Chapter 8, the implementation and evaluation of the proposed solutions presented in Chapter 5, 6, and 7 will be described. Experimental results will also be analysed and discussed in details. Finally, chapter 9 summarises the research and outlines some suggested future work.

Chapter 2

IEEE802.11 Wireless LANs

The IEEE 802.11 wireless local area networks (WLANs) have become a wide spread wireless access technology and a flexible alternative to the traditional wired access. WLAN technologies use radio frequencies as the medium of transmission, and thus eliminate cables and wiring for simpler network management at effective costs. This chapter starts with a brief background on the WLAN architecture, after which some related standards and security issues are discussed. An in-depth study on the IEEE 802.11i in an enterprise environment will also be given. The material presented in this chapter provides the necessary background and building blocks, which will help explain further details throughout the thesis.

2.1 WLAN Architecture Overview

The IEEE 802.11 defines two WLAN topologies: the Ad-hoc and the Infrastructure mode. The ad-hoc mode is also referred to as Independent Basic Service Set (IBSS). It is a peer-to-peer network in which no dedicated entity is required to assume the role of a gateway router. Enterprise WLAN normally works in infrastructure mode. Thus the subsequent chapters will refer to the infrastructure mode WLAN.

The IEEE 802.11 infrastructure mode architecture is comprised of several components that interact to provide connectivity and mobility. The main components are as follows:

Station (STA). A STA is a wireless endpoint device, which enables end users to gain access and utilise resources provided by the wireless network. Such a device provides 802.11 functionality and implements 802.11 medium access control (MAC) and physical layer (PHY). Examples include laptop computers, PDAs, mobile phones and

other consumer electronic devices with WiFi (IEEE 802.11) capabilities. In the thesis, the word “client” and “STA” are used interchangeably to refer to the same thing.

Access Point (AP). The WLAN Access Point (AP) is a STA with additional functions to support bridging (i.e., layer 2 forwarding) and 802.11 operation management. An AP is usually connected to a wired network, such as a corporate LAN, and relays data between the wireless network and the wired network. An AP provides user mobility by allowing a STA to move freely within the AP’s coverage range while maintaining connectivity to the AP.

Basic Service Set (BSS). A BSS basically represents the coverage area of an individual AP. It consists of a group of STAs which communicates directly to the AP. All the frames exchanged between STAs are relayed by the AP. A BSS is the smallest building block of a WLAN in which clients execute the same MAC protocol and compete for access to the same shared wireless medium. Each BSS is identified by a BSS identifier (BSSID) of a 6-byte length. In the infrastructure mode, the BSSID is actually the AP’s MAC address.

Distribution System (DS). A typical enterprise WLAN requires several APs to provide wireless access over a larger area. To enable roaming between multiple APs while maintaining connections to wired network resources, the 802.11 standard specifies a DS, which provides wired interconnections between APs.

Extended Service Set (ESS). The 802.11 allows several BSS's to be linked together via a DS, to form an ESS. An ESS is identified with an ESSID (Extended Service Set Identifier), which is a 32-character identifier in ASCII format. The ESSID (usually shortened to SSID) represents the name of the wireless network, and in a way acts a first-level security measure, since it is required for a STA to know the network SSID in order to connect to it.

Within an ESS, if all BSSs use the same SSID, STAs can roam between APs without much network disruption. The STAs can discover nearby APs by scanning and change its association when its location is changed (detected by measuring the signal strength). On the other hand, if BSSs use different SSIDs, a roaming STA has to change its SSID

to match the one used by the new BSS in order to associate to it. This usually requires a hardware reset, and thus, causes longer roaming delay. Figure 1 shows the structure of an infrastructure mode WLAN.

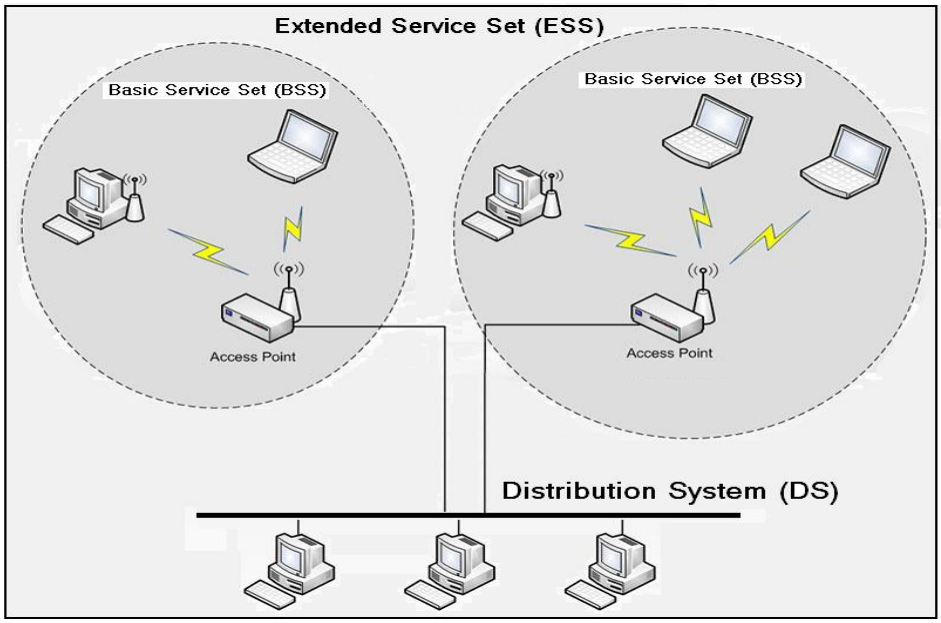

Figure 1: A basic structure of an infrastructure mode WLAN

In this research, it is assumed that all the APs in the ESS use the same SSID. This is what most enterprise networks actually do in practice, as it significantly reduces handoff delays.

2.1.1 IEEE 802.11 Management Frames

The IEEE 802.11 management frames enable STAs to establish and maintain connectivity to the WLAN. The following are the commonly used management frames:

- **Beacon frame:** An AP transmits beacon frames at a regular interval to advertise its existence. Those beacon frames contain useful information about the capabilities of the AP and the network identifier.

- **Probe request frame:** A STA can actively send a probe request frame to search for the existence of APs within reach.

- **Probe response frame:** If an AP receives a probe request with compatible parameters such as the SSID and the data rates, it can reply with a probe response frame to allow the STA to join the AP.

- **Authentication frame:** Before a STA can establish a connection with an AP, it must authenticate itself to the AP. This frame is used to exchange the required authentication information.

- **De-authentication frame:** The frame is used to change the STA's state from authenticated to unauthenticated. A STA needs to repeat the authentication process again to become authenticated after being de-authenticated.

- **Association request/response frame:** Once a STA has authenticated to the AP, the association request is used to establish an association. The process of association checks that the capabilities of the STA and the AP are compatible, and negotiates some parameters, such as the data rate. The AP replies to the STA with an association response to notify whether the association is successful. For each successful association, the AP reserves memory to store some state information and an association ID for the associated STA.

- **Disassociation frame:** The frame is used change the STA's state from associated to unassociated. The STA needs to repeat the association process to become associated again after becoming unassociated.

- **Re-association request/response frame:** When a STA moves from one AP to another within the same ESS, the re-association request is used to join the new AP. The re-association response is similar to the association response frame, which contains information regarding the association relationship and an association ID. The only difference is that this frame is used as a reply to re-association requests.

2.1.2 Getting Connected in IEEE 802.11

In the infrastructure mode, a STA needs to go through three stages in order to establish connectivity with an AP. The three stages in sequence are: discovery, authentication, and association. Each of them will be briefly described below in this section. Figure 2 summarises the connection establishment process and the state transitions with different management frames.

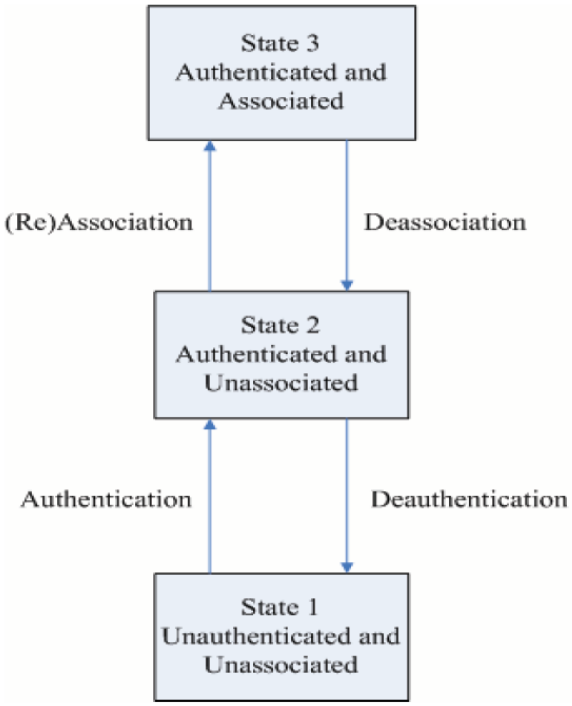

Figure 2: 802.11 state transitions and related management frames

1. Discovery

The purpose of the discovery stage is to scan and search for existing APs within reach. The scan can be passive or active. For passive scan, the STA identifies nearby APs by listening for beacons that are periodically broadcasted by APs. In contrast, active scan requires the STA actively broadcasts probe request messages. If any AP receives the probe request, it immediately replies with a probe response, which contains similar information in the beacon frame. This allows a STA to learn about the APs within reach more quickly.

By the standard default, beacons are broadcasted 10 times every second. This might seem to be plenty for a STA to find a right AP. In an enterprise network environment, however, there can be multiple frequency channels. For a complete scan, a STA needs to go to each channel and wait for 100ms for a beacon. Thus, this process incurs a long delay. If a connected STA is required to find a new AP due to the weak signal strength, performing a passive scan can cause application service disruption because of the long scan delay. Therefore, most implementations, including the testbed implementation in this research, use active scan in order to minimise the handoff delay.

2. Authentication

Authentication is the process by which a STA proves its identity to be eligible to join the network. The IEEE 802.11 standard has defined two types of authentication: Open System and Shared Key authentication. They are link-layer (Layer 2) authentication schemes and were defined in the early version of the standard back in 1997. With the Open System authentication, two messages are exchanged. The first one asserts identity and requests the identity of the other communicating party. The second message returns the result of the authentication, which can be success or failure. Open System authentication is equivalent to trusting everyone, thus does not provide any security.

Shared key authentication intends to provide security by authenticating those clients who know a shared secret key. This scheme assumes that the secret key is delivered independently of the authentication process through a separate secure channel. Using cryptography, a challenge text is encrypted at one station, transmitted, and decrypted at the other station. If the result is identical with the challenge text, the stations are proved to have the same key, and successfully authenticated. However, this process fails to be a secure authentication scheme because all the information required to reconstruct the key are made public. Hence, Shared Key authentication, much like Open System authentication, does not provide much security. Because of the weak security, when IEEE 802.11i was later proposed, the authentication is done with the IEEE 802.1X port based authentication instead of the Shared Key authentication. A detailed description of the IEEE 802.1X authentication will be presented in Section 2.3.1.

3. Association

After the authentication the STA has obtained permission to connect to the AP, and the association is the last step before the connection is established. In the 802.11 standard the concept of "connection" is referred to as association. When a STA is associated with an AP, it is eligible to send and receive data from the network[1]. The STA sends an association request and the AP replies with an association response indicating a successful connection. After this stage, data sent from the STA is forwarded onto the DS to which the AP is connected. Similarly, data from the DS intended for delivery to the STA is forwarded by the AP. As a part of the association process, for each associated STA the AP generates and assigns an association ID (AID), which is an arbitrary number used to identify the association relation between the STA and the AP.

2.2 WLAN Security Standards

A primary concern in wireless access provision is network security. Confidential information can travel over the air in clear text if no additional data protection is used. In order to bring data privacy and security in line with wired LANs, the 802.11 standard defined a security protocol called Wired Equivalent Privacy (WEP).

To prevent eavesdropping activities, WEP makes use of the RC4 (Rivest Cipher) algorithm to encrypt and decrypt 802.11 packets with a 40-bit symmetry key[2]. To add randomness to the key encryption, an initialization vector (IV) was added to the fixed-

[1] This is with regard to the original 802.11 standard. Being associated gives the STA full network access straight away. However for IEEE 802.11i, as presented in section 2.3, association only allows the STA to begin the full authentication process (achieved through IEEE 802.1X) required for secure network access

[2] A later revision enables the use of 104-bit key. However, this is still not secure.

length encryption key. However, it was shown in [10] that the IV sequence is too short and therefore repeats within a timeframe that allows the calculation of the key.

Another issue with WEP is that it does not include any key management protocols, so the pre-shared key must be manually configured into devices. If the shared key is leaked, the WEP security is compromised. Because the key is not likely to be changed frequently, this further increases the risk of key being sniffed or cracked.

Furthermore, cryptoanalysts have identified vulnerabilities in the RC4's key scheduling algorithm that could be exploited by hackers [11]. Walker in [3] also showed that it is infeasible to achieve privacy using WEP encapsulation, even with an expanded key size. This work presented an attack against WEP and demonstrated that the attack can always succeed regardless of the key size or the cipher used. Based on those vulnerabilities, there exist several attack tools such as [12, 13] that enable hackers to easily crack the WEP key and gain access to the WLAN.

To provide a stronger security alternative to WEP, in 2001 the Wi-Fi Alliance introduced an enhanced security scheme called Wi-Fi Protected Access (WPA), a new industry standard for WLAN security. Later in 2004, IEEE also finalised the 802.11i standard, also known as Robust Security Network (RSN) or WPA2. One significant difference between these new standards and WEP is that the new standards separate the user authentication from the message protection process [14].

This section begins by looking at WPA and the transition into the 802.11i RSN. Section 2.3 will provide the background needed to understand the authentication process in IEEE 802.11i. Section 2.4 will describe the 802.11i key management scheme and related standards that comprise the RSN.

2.2.1 WPA

While the security issues of WEP were widespread and the IEEE 802.11i security enhancements were still under development, WPA was proposed as an intermediary

solution. Legacy products only require a firmware upgrade to provide WPA, as apposed to IEEE 802.11i which requires hardware replacement.

WPA offers standards-based and interoperable security enhancements that eliminate most 802.11 security issues and greatly increase the level of data protection and access control for WLANs. As a temporary solution and to be compatible with existing products, WPA replaces WEP with the Temporal Key Integrity Protocol (TKIP), which is a compromise between strong security and backward compatibility. It continues the use of RC4 as the encryption algorithm, however, an additional keyed packet authentication mechanism, Michael, is included to protect against replay attacks.

To provide access control and key management, WPA offers two authentication modes: Personal mode and Enterprise mode. The WPA Personal mode (WPA-PSK) uses pre-shared key for data encryption. It does not provide key management and requires no additional servers. On the other hand, the enterprise mode integrates the IEEE 802.1X to support upper-layer authentication with an external authentication server (e.g., RADIUS). A new method of key exchange process called the four-way handshake is used for generating and exchanging data encryption keys between the AP and the STA.

When a STA wants to join the RSN, it still performs the IEEE 802.11 authentication with an AP. The Open System authentication is used here, so there is no security at this stage. After STA is associated to the AP, the real authentication work takes place over the IEEE 802.1X with an upper layer authentication method, such as the EAP-TLS. The authentication server will verify the STA's credentials and makes access control decisions. Once the STA is authorised by the server, data can then be exchanged between the STA and the network. More detailed description of the IEEE 802.1X will be presented in Section 2.3.1. The four-way handshake key exchange will be explained in detail later in Section 2.4.3.

In summary, WEP enhanced the WLAN security by providing dynamic encryption key, authentication and replay detection. However, weakness is predestined due to the limitation imposed by its re-use of legacy hardware [15].

2.2.2 IEEE 802.11i

To further enhance the WLAN security, IEEE started developing a much stronger security scheme called the IEEE 802.11i [16] standard which was released in June 2004. Utilising the power of hardware-based crypto-computations, the Advanced Encryption Standard (AES) is used instead of RC4 for data encryption. Further, by replacing TKIP with CCMP (Counter Mode CBC-MAC Protocol), 802.11i enables better management of dynamic keys as well as better data confidentiality. To be compatible with existing products, the 802.11i allows an alternative option to use TKIP for confidentiality and Message Integrity Code (MIC) for data integrity [17]. Though TKIP does offer a much better security than WEP, CCMP is the default and recommended protocol in 802.11i due to its arguably uncompromised confidentiality and integrity. However, 802.11i has not been designed to address potential threats to availability. The management frames are still unprotected and unauthenticated. Thus, 802.11i is still susceptible to DoS attacks.

Table 1 shows a summarised comparison of related encryption properties between WEP, WPA, and IEEE 802.11i standards.

	WEP	**WPA**	**IEEE 802.11i**
Encryption Algorithm	RC4	RC4	AES
Key Length	40/104 bits	128 bits	128 bits
IV Length	24 bits	48 bits	48 bits
Integrity Check	CRC-32	Michael	CCM
Key Management	Nonce	TKIP	CCMP

Table 1: WEP/WPA/802.11i Comparison

2.3 IEEE 802.11i Security Management

Security in the WLAN can be broken down into three components: authentication framework, authentication methods, and encryption. A robust authentication framework is something that was missing from WEP and contributed to its demise. The original 802.11 standard provides link layer authentication with the Open System or Shared Key authentication, as described in section 2.1.2. The IEEE 802.11i provides authentication by utilising an additional separate protocol called IEEE802.1X, which is a reliable framework for access control that leverages EAP (Extensible Authentication Protocol) to provide centralised and mutual authentication.

This section provides a brief introduction to the IEEE 802.1X and describes the EAP and RADIUS protocols which are used to complement the 802.1X authentication in establishing a Robust Security Network Association (RSNA). More details on the RSNA and the full establishment procedure will be presented in section 2.4.

2.3.1 Access Control with IEEE802.1X

802.1X is a port-based authentication protocol used for both wired and wireless networks. It is based on the EAP (Extensible Authentication Protocol) [18] to provide compatible authentication and authorisation mechanisms for devices connecting to the RSN. The purpose of the 802.1X is to implement access control at the point at which a STA joins the network. The access control is done by blocking access to network resources until a STA is properly authenticated to an AP. The 802.1X framework defines the following three entities:

- **Supplicant**: The software that implements the client side of the 802.1X standard. The Supplicant is loaded onto the client's device and is used to request network access. In the context of the thesis, it refers to a wireless client or a STA.

- **Authenticator**: A device that sits between the supplicant that needs to be authenticated and the external authentication server that makes authentication decisions. In the context of the thesis, it refers to a wireless AP.

- **Authentication Server (AS)**: A backend server which resides in the DS that participates in the authentication of all supplicants. It receives authentication requests and makes authentication decisions based on supplicant's credentials. It refers to a RADIUS server in the context of the thesis.

802.1X grants per-port access to supplicants requesting access to network resources. For each supplicant, the authenticator would associate two logical ports: uncontrolled port and controlled port. The uncontrolled port is only used for the exchange of authentication messages. A controlled port is an entry point to the network resources. Only authorised STAs can pass traffic through the controlled port. Hence, until a client is authenticated by the authentication server, the network access is blocked by only allowing communications to the uncontrolled port.

Both the supplicant and the authenticator have a PAE (Port Access Entity) that manages and executes authentication related algorithms and mechanisms. Figure 3 below depicts the 802.1X framework.

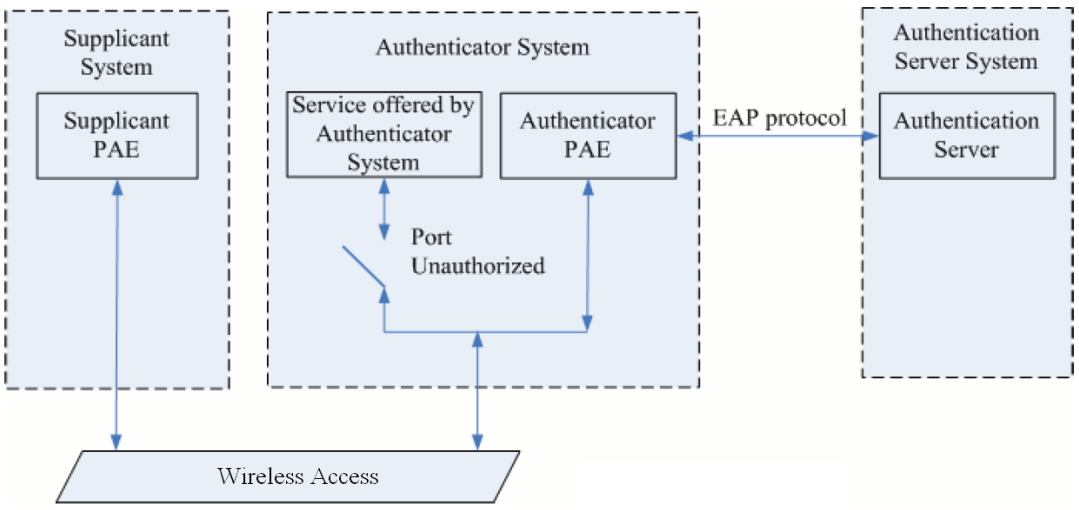

Figure 3: IEEE 802.1X Access Control Framework

2.3.2 EAP

One of the major advantages of using 802.1X in RSN is that a variety of upper layer authentication methods can be used. EAP is an envelope that can carry many different kinds of authentication methods including, challenge/response, one time passwords, digital certificates, etc. EAP defines four basic packet types: request, response, success, and failure. A basic EAP dialog is depicted in Figure 4.

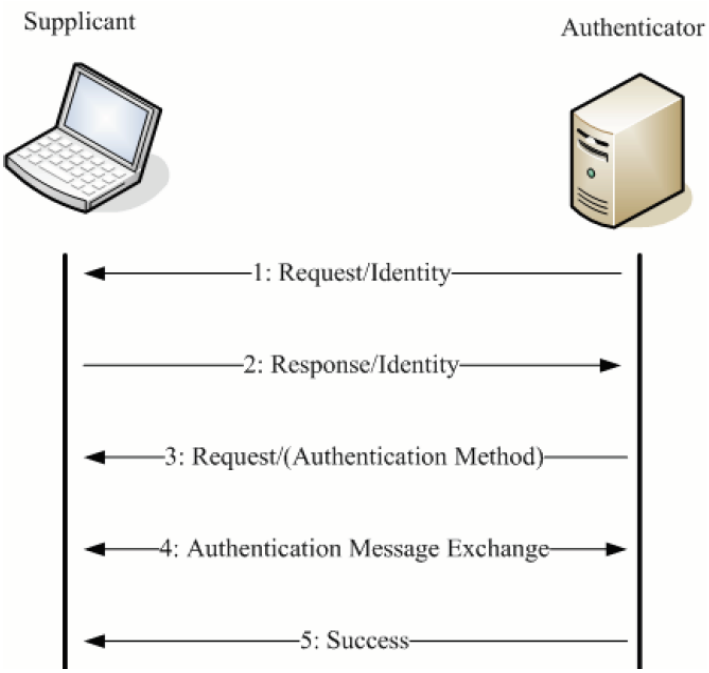

Figure 4: EAP Message Exchange Procedure

EAP communication involves the bidirectional transmission of request and response frames, finally ending with a success or failure notice. The process starts with the message **Request/Identity**, sent by the authenticator to a new supplicant. The supplicant replies with the message **Response/Identity** containing its identifier that will be understood by the authentication server. This is then followed by an upper-layer authentication exchange, and finally a **Success** or **Failure** notification will be sent to the supplicant to indicate the result of the upper-layer authentication.

Within EAP, success and failure notification frames are neither acknowledged nor integrity protected. Although results themselves are not protected, using an authentication method (e.g., TLS) that provides integrity protection and replay

protection can enhance the security. Hence, the testbed implementation and analysis in this research will be based on the EAP-TLS authentication method. However, DoS attacks are still possible. More DoS vulnerability issues are discussed in Chapter 3.

2.3.3 EAPoL

EAP was originally designed to independently transport authentication exchange messages used by upper layer authentication methods with dial-up connection via a modem. Before 802.1X became prominent in WLAN authentication, EAP was mainly used to authenticate dial-up users. In order for EAP authentication messages to be transported on a LAN, they need to be encapsulated. IEEE 802.1X defined EAP over LAN (EAPoL) to encapsulate EAP packets by prefixing an Ethernet header onto EAP messages so that they can be transmitted over Ethernet. EAPoL defines the following four types of messages that are used by RSN:

- **EAPOL-START:** When the supplicant is connected to the network and ready to authenticate, it can send this message to the authenticator to initiate the authentication procedure.

- **EAPOL-KEY:** With this message type, the authentication can send encryption keys to the supplicant. To improve security, IEEE 802.11i modified the original format in some ways to allow RSN to establish encryption keys and also to validate that both supplicant and authenticator have correct keys before granting access.

- **EAPOL-PACKET:** This frame is used to encapsulate the actual EAP messages in order to transport them on LAN.

- **EAPOL-LOGOFF:** This message type is used when a supplicant wants to be disconnected from the network.

2.3.4 RADIUS

RADIUS (Remote Authentication Dial-In User Service), as defined in RFC 2865 [19], is a networking protocol that provides centralised authentication, authorization, and accounting (AAA) for devices to access services in an IP based network. RADIUS is not specifically part of the IEEE 802.11i standard, but most of the existing enterprise WLAN implementations use it for transferring authentication messages between the authenticator and the authentication server. A RADIUS server is an authentication server that supports RADIUS capabilities, and is also the entity that decides the access permission of a supplicant.

In this research, RADIUS is used to provide the AAA service. Although Diameter [20], a successor to RADIUS, can also be used for the same purpose, the authentication concept and operations are very similar. The main difference is that RADIUS uses UDP as the transport layer while DIAMETER uses SCTP or TCP to provide more reliable data delivery.

2.3.5 Putting All Together: 802.1X in RSN

The 802.1X framework allows the secure exchange of client credentials, and prevents virtually any unauthorised network access due to the fact that authentication is done before a network IP address is assigned[3]. The framework consolidates decision-making at the RADIUS server, so that passwords no longer have to be individually configured into every STAs or APs. It also allows STAs to authenticate with credentials other than MAC address, which are easy to spoof. Supplicant's credentials are securely passed to the authenticator via a secure EAP method (e.g., EAP-TLS), which are then forwarded to the authentication server via EAP in RADIUS. This is a Layer 3 transmission that allows for the secure passing of authentication messages (authentication request,

[3] This is because 802.1X operates at Layer 2 - the Data Link layer.

authentication result) and access authorization (accept, reject) between the Authenticator and Authentication Server. Figure 5 below outlines the 802.1X authentication and authorization process in the RSN environment.

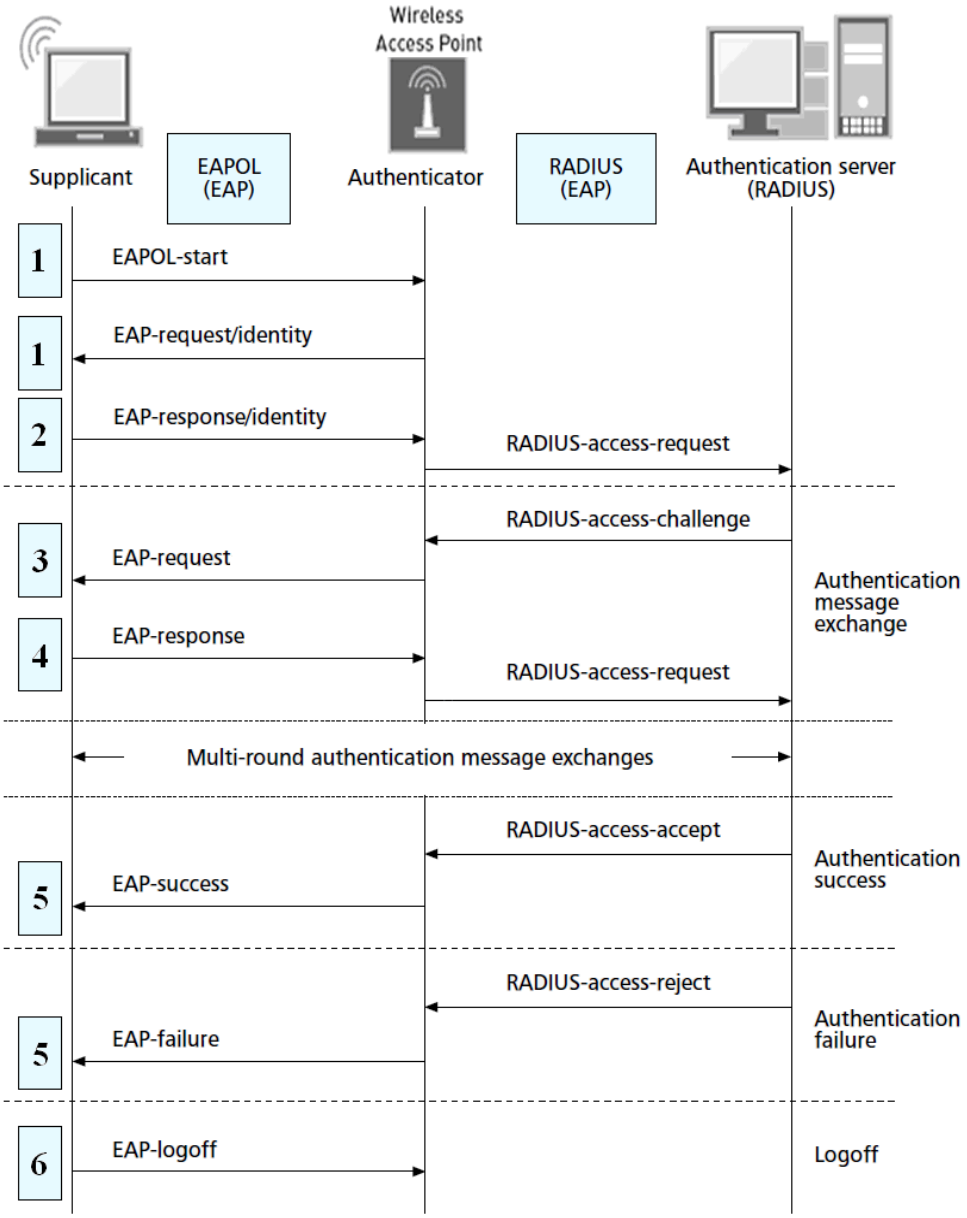

Figure 5: 802.1X Authentication Framework in RSN

1. Either the authenticator or the supplicant can initiate an 802.1X authentication process. The authenticator initiates by sending an EAP-Request/Identity frame. The supplicant initiates an authentication message exchange by sending an

EAPOL-Start frame, to which the authenticator responds with an EAP-Request/Identity packet.

2. The supplicant sends an EAP-Response/Identity message to the authentication server via the authenticator to provide its identity. The authenticator encapsulates the EAP-Response/Identity message in RADIUS-Access-Request for the forwarding to the server.

3. The authentication server challenges the supplicant by sending a RADIUS-Access-Challenge to the authenticator, which then forwards it in the form of an EAP-Request frame to the supplicant.

4. The supplicant provides its authentication credentials (e.g., username and password) in an EAP-Response message, which is again forwarded to the authentication server by the authenticator in the format of RADIUS-Access-Request.

5. After several message exchanges for the chosen EAP authentication method (e.g., EAP-TLS), the authentication server determines the access permission and either sends an RADIUS-Access-Accept or RADIUS-Access-Reject to the authenticator. On receipt of RADIUS-Access-Accept, the authenticator changes the supplicant's state to Authenticated, unblocks the controlled port (allowing full access), and sends an EAP-Success message to indicate the success of authentication to the supplicant. Similarly, EAP-Failure is sent instead if the authentication fails.

6. To disconnect from the network, the supplicant can send an EAPOL-Logoff frame to the authenticator. This causes the authenticator to change the supplicant's state to unauthenticated and the controlled port to unauthorised, thus blocking the network access.

2.4 RSNA and Key Management

The IEEE 802.11i introduces the concept of a Robust Security Network (RSN). A RSN is defined as a wireless security network that only allows the creation of Robust Security Network Associations (RSNA). An RSNA is a logical connection between communicating entities established through the 802.11i key management scheme. A successful 802.1X authentication results in both the supplicant and the authenticator holding a pairwise master key (PMK). The four-way handshake process validates that both entities share the same PMK, synchronises the installation of encryption keys, and confirms the selection and configuration of data confidentiality and integrity protocols.

The thesis will investigate the RSNA and address its DoS vulnerabilities. First, components that are involved in a RSNA establishment are described in the following sections. Chapter 3 will then analyse the RSNA procedure and identifies potential vulnerabilities to DoS attacks.

2.4.1 RSNA Procedure

The 802.11i RSNA establishment procedure makes use of 802.1X authentication and key management protocols. The complete handshakes for establishing a RSNA are shown in Figure 6 which involves the following six stages:

Stage 1: Network Discovery

Stage 2: 802.11 Authentication and Association

Stage 3: EAP/802.1X/RADIUS Authentication

Stage 4: four-way Handshake

Stage 5: Group Key Handshake

Stage 6: Secure Data Communication

The procedure starts with the network discovery using probe messages, followed by 802.11 authentication and association, as described in Section 2.1.2. Because RSNA

relies on 802.1X for achieving mutual authentication via upper-layer authentication methods, Open System authentication is used at this stage.

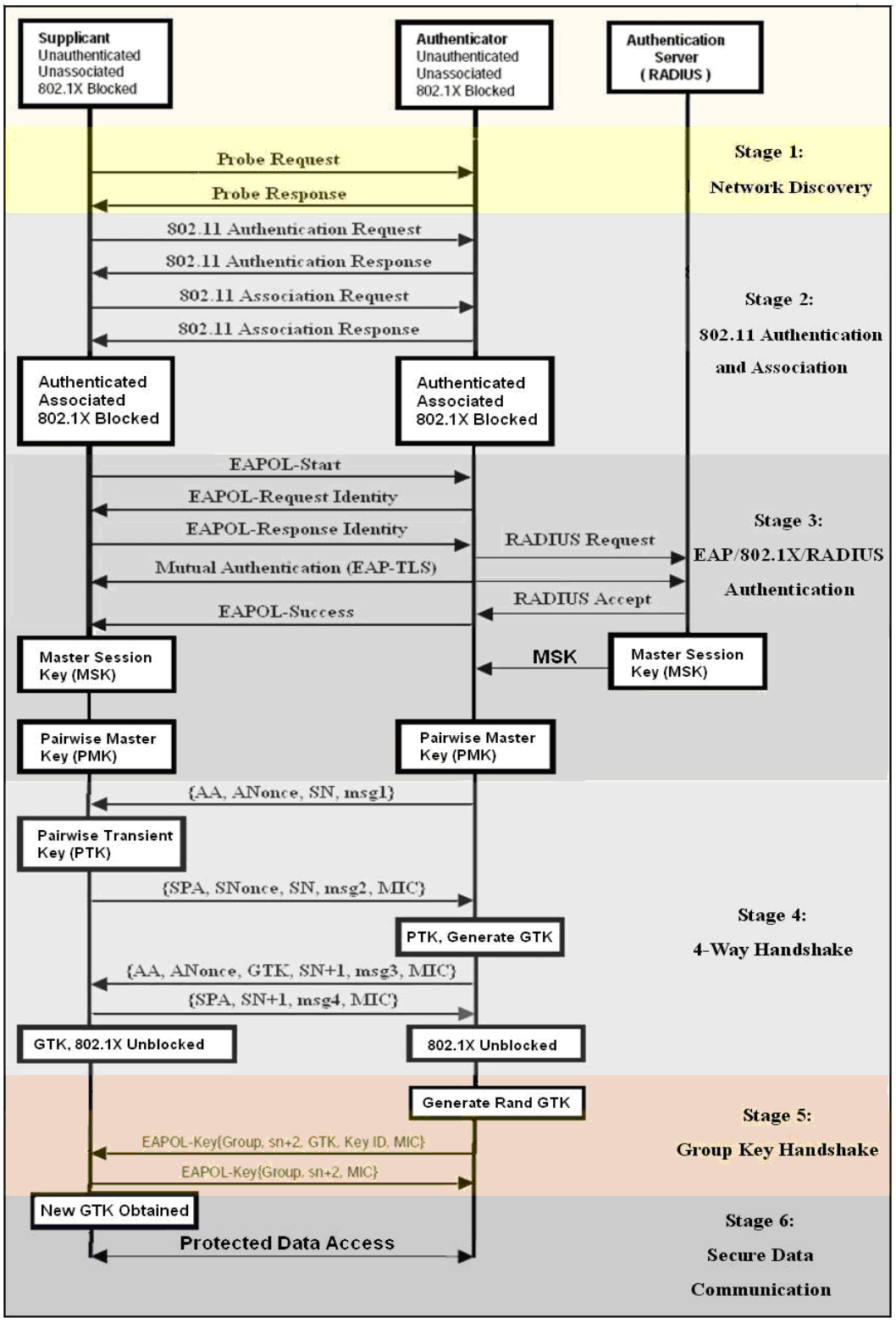

Figure 6: RSNA Establishment Procedure

After the STA has associated to the AP, no access to the network is allowed yet because the 802.1X controlled port still remains blocked. To unblock the port, the 802.1X authentication procedure (Stage 3), as described in Section 2.3.5, takes place to allow the STA to authenticate to the remote authentication server. If the authentication is successful, both the STA and AP will independently generate a common secret, called the Master Session Key (MSK). The STA then uses this MSK to derive a Pairwise Master Key (PMK). On the server side, the MSK is securely transferred to the AP via the RADIUS protocol. The AP then uses the MSK to derive a matching PMK.

In Stage 4, the four-way handshake is performed to confirm the liveness of the STA and the freshness of the PMK. During the handshake process, further crypto-computation uses the PMK to generate a Pairwise Transient Key (PTK), which will be used to encrypt unicast packets for providing confidentiality and integrity. To protect multicast packets, Stage 5 performs the group key handshake to generate a Group Transient Key (GTK) for encrypting multicast packets. Finally in Stage 6, a secure data communication is established between the AP and the STA, since all the data packets are encrypted and protected using those transient keys.

The following sections will describe the RSN key hierarchy and the four-way handshake in more details.

2.4.2 RSN Key Hierarchy

RSN defines two types of keys: **pairwise key** and **group key**. The pairwise key is used to protect unicast data sent between a STA and the AP. Hence, each STA needs to store one pairwise key, and the AP needs a set of pairwise keys (one for each STA that is associated). The group key, on the other hand, protects multicast (including broadcast) data that are received by multiple STAs who form a trusted group. Hence, a group key is shared by all the members of a multicast group. Both pairwise key and group key are managed in the form of a key hierarchy.

2.4.2.1 The Pairwise Key Hierarchy

For the pairwise key hierarchy, the PMK, which is delivered from the upper-layer authentication protocol used during the 802.1X authorization, is at the top level of the hierarchy. The PMK (which is supposed to be a secret) is not directly used to encrypt packets, but used to further derive a set of four transient keys that are actually used for data encryption and integrity checks. The four transient keys together are called the Pairwise Transient Key (PTK).

The transient keys are used to provide two layers of protection – the EAPoL handshake messages and the user's data packets. For each layer, two cryptographic functions are used: encryption and integrity. The four transient keys are:

- Data Encryption Key (128 bits)
- Data Integrity Key (128 bits)
- EAPOL-Key Encryption Key (128 bits)
- EAPOL-Key Integrity Key (128 bits)

The first two keys provide encryption and integrity of user's unicast data packets; similarly the latter two keys provide encryption and integrity to the four-way handshake messages. Theses four keys are referred to as the transient keys because they are refreshed (re-computed) every time a STA associates or re-associates to an AP.

Because the transient keys need to be refreshed, providing randomness to the creation of the transient keys is required. This is achieved by including a couple of special random values called "nonces" in the computation. Both the STA and the AP generate and exchange their own nonce, and then derive the transient keys by including both nonces in the computation. In order to bind the two entities' identity into the keys, the MAC address of each is also included in the computation. Figure 7 summarises the computation process and shows the inputs to the computation and the four output keys.

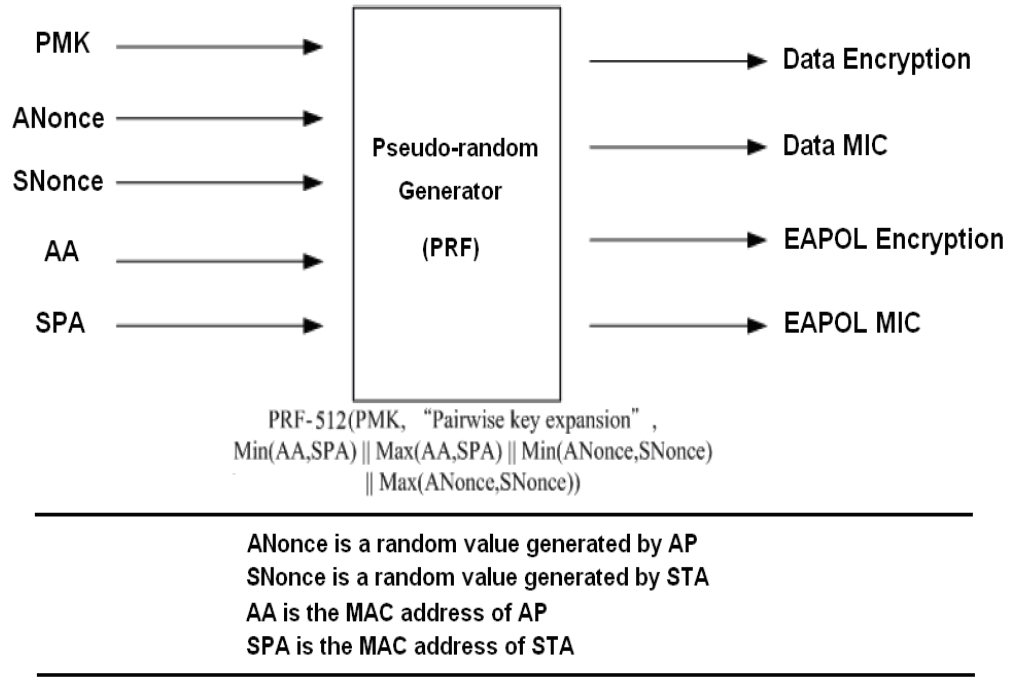

Figure 7: Computation of the Transient Keys

RSN defines a pseudorandom function (PRF-X) to convert a set of variable length input strings into an output string of X bits. Based on this function, the computation of the pairwise transient keys are derived. Figure 8 summarises the computation of transient keys for TKIP encryption. For AES encryption, the same key is used for both data encryption and data integrity.

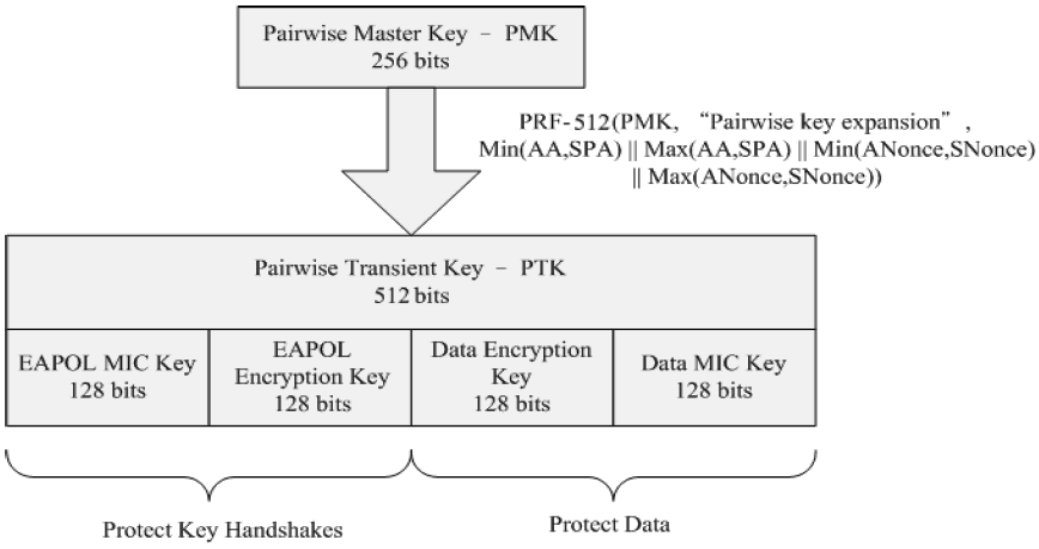

Figure 8: RSN Pairwise Key Hierarchy

2.4.2.2 The Group Key Hierarchy

To allow multicast data to be received by multiple STAs which form a trusted group, a group key needs to be shared by all the members of the group. The distribution of the group key is done using EAPOL-Key messages, as is for pairwise keys. The AP first creates a 256-bit group master key (GMK), and then derives the 256-bit group transient key (GTK) by combining with a nonce value and the MAC address of the access point. The GTK contains the following two group transient keys:

- Group Encryption Key (128 bits)
- Group Integrity Key (128 bits)

Via the pairwise secure communication that has previously been established, the AP can safely send the derived GTK to the STA.

2.4.3 Four-Way Handshake

The previous sections have described the mutual authentication process via 802.1X, and outline the key derivation using PMK. To allow a STA to verify the identity of the AP, the AP is required to prove its possession of the PMK. In RAN, the process of proving PMK possession is combined with the process of computing pairwise transient keys. There are four steps required, referred to as the four-way handshake, and the exchange is done using EAPOL-Key messages. This section describes the four-way handshake in more details.

The four-way handshake serves the following purposes:

- To confirm that both STA and AP possess a current PMK,
- To confirm the cipher suite selection and derive a fresh PTK from the PMK,
- To synchronously install the encryption and integrity keys as well as the GTK into both entities.

During the four-way handshake, four EAPOL-Key frames are exchanged between the STA and the AP as depicted in Figure 9. The AP initiates the process by sending Message 1, including a random number (ANonce). This message is solely used to send AP's nonce to the STA. At this starting state, no transient keys are known so the MIC cannot be computed. Therefore, Message 1 is completely unprotected because it is neither encrypted or having a MIC.

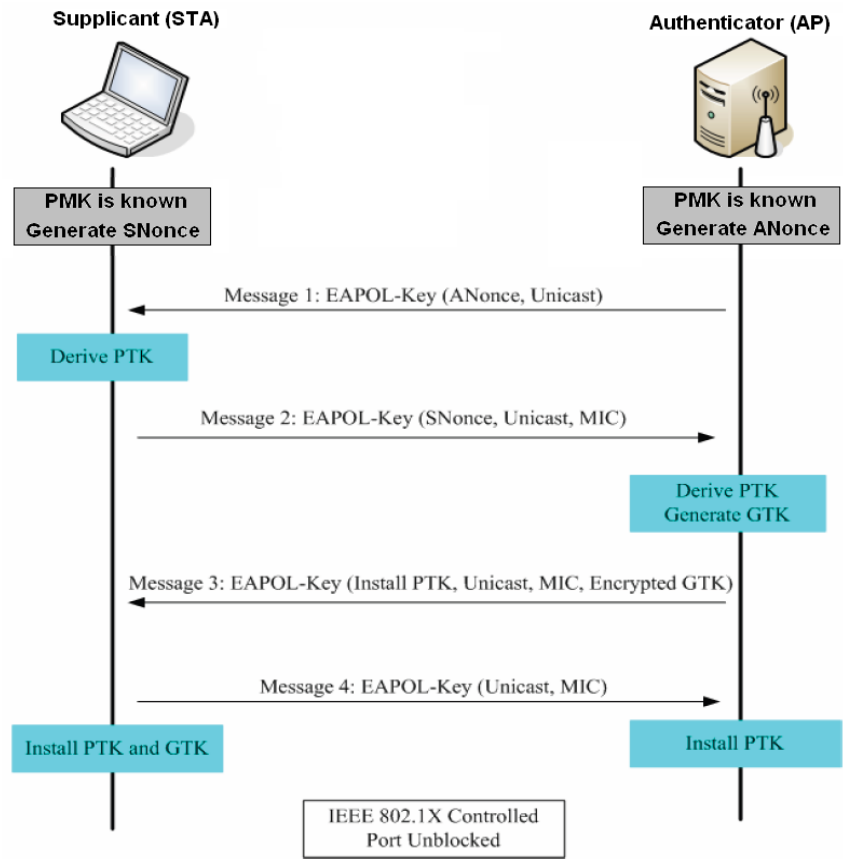

Figure 9: The four-way handshake exchange

After generating its own SNonce and extracting the ANonce from the previous message, the STA is now able to compute the PTK from the PMK, using the formula shown in Figure 7. Once the STA has computed the PTK, it sends Message 2, containing its SNonce, to the AP. This message includes a MIC value (computed with the EAPOL-Key Integrity Key) and thus proves to AP that the STA actually possesses a valid PMK.

When the AP receives this integrity protected Message 2, it is able to extract the enclosed SNonce (because the message is unencrypted) along with the STA's RSN Information Element (RSN IE) which confirms the cipher suite selection. The AP is then not only able to derive the same PTK on its side, but also can verify that the STA is in possession of the current PMK and has derived the transient keys properly. The AP also uses the derived EAPOL-Key Integrity Key to verify the MIC of Message 2, to ensure that the message has not been tampered. If the MIC validation fails, the AP will drop this message. If these checks are successful, the AP also further generates a GTK that will be used for encrypting multicast packets.

The remaining two handshake messages are used to ensure that keys are put into effect in a synchronised fashion. The AP sends Message 3 to the AP, which contains an "Install PTK" instruction. To also inform the STA about the group key, the GTK is first encrypted with the EAPOL-Key Encryption Key and the result is put into this message. Again, this message is integrity protected with the EAPOL-Key Integrity Key. When the STA receives this message, the MIC is first validated, and then it decrypts to obtain the GTK.

The transmission of the fourth and last frame allows the STA to announce that the PTK and GTK will be installed. At this point in time, both entities have proved their knowledge of the previously negotiated PMK to each other and derived the transient keys required to protect subsequent data packets. After the successful completion of the four-Way Handshake, the STA is granted access to the network resources via the 802.1X controlled port of the AP. Both entities will also start using encryption to protect unicast and multicast packets.

2.4.4 Group Key Handshake

The 802.11i standardizes a process for group keys to be periodically renewed in order to expire the key material of leaving STAs. By means of a two-way exchange of integrity protected EAPOL-Key messages, the AP and the associated STAs can negotiate a new GTK in a secure manner.

To renew the group key, the AP simply derives a new GTK, encrypts it with the EAPOL Encryption Key that was created as part of the pairwise key handshake, and passes it to any affected STA which in turn acknowledges the receipt by a subsequent EAPOL-Key message.

2.5 Chapter Summary

The various confidentiality and integrity vulnerabilities found in WEP have been properly addressed by the new security standard IEEE 802.11i. This enhanced standard provides a security framework utilising several useful and approved protocols together to deliver robust protection for WLANs. Combining the enhanced user authentication, strong confidentiality with new underlying cipher, and a reliable integrity verification scheme, IEEE 802.11i has pushed WLAN security to a much higher level comparable with the security in wired networks. However, the major focus of the standard is to secure higher layer user and protocol information (i.e., the 802.11 data frames), link layer security is not addressed. Hence, the unprotected management frames are vulnerable to a wide spectrum of known attacks. Moreover, the protocol execution of IEEE 802.11i/IEEE 802.1X also has a few weaknesses that can be exploited to launch DoS attacks. Chapter 3 will start by identifying those vulnerabilities in the current 802.11i standard and look into DoS attacks that exploit those link layer vulnerabilities in more details.

Chapter 3

DoS Vulnerabilities in WLAN

The goal of Denial-of-Service (DoS) attacks is to compromise the availability of a system so that legitimate users are prevented from accessing network resources and services. Any activities that are blocking network access, causing excessive delays, consuming valuable network resources, etc, are considered as DoS attacks.

Several prior research works have identified numerous DoS vulnerabilities in a WLAN from the physical Layer to the application layer [8, 12, 24, 42]. Compared to DoS attacks in the wired networks, WLAN DoS vulnerabilities appear to be more threatening for several reasons. First, an attacker can launch an 802.11i attack easier with only moderate equipment. Second, it is a lot more difficult to detect and mitigate wireless DoS attacks. Although prevention measures at network and physical layers have been extensively dealt with in many research works [21-24], MAC layer DoS vulnerabilities are not fully addressed. The main reason seems to be that the frequency jamming vulnerability has been a well known indigenous issue in WLAN, providing availability protection at link layer is often considered as unimportant. However, it is arguable that there is a significant difference between physical layer attacks targeting channel capacity to deny any communication, and link layer attacks disrupting the services provided by APs and the connection states of STAs.

Although the 802.11i standard has improved WLAN security with data confidentiality, integrity, and mutual authentication, link layer network availability is still weak because it was not an original design objective. As a result, the most widely exploited and the most threatening vulnerabilities arise from those unprotected management frames. For example, an attacker can easily launch a DoS attack against a specific STA or the entire BSS by flooding forged deauthentication or disassociation frames to

terminate existing connections. Those DoS attacks have been major obstacles in providing continuous service availability and achieving required network performance.

This chapter describes various flooding-based DoS attacks that exploit the vulnerabilities of unprotected WLAN management frames. The scope of analysis will focus on DoS vulnerabilities at the MAC Layer. Other possible DoS attacks in PHY or upper layers are outside the scope of this research. Later in Chapter 4, the impact of DoS attacks will be analysed and evaluated in a testbed environment.

3.1 802.11i State Machine Vulnerability

IEEE 802.11i defines a state machine that the STA implements for tracking authentication and association states. The AP also maintains a separate state machine for each of its authenticated/associated STAs. Figure 10 adopted from [14] depicts the state transitions and the controlling management frames.

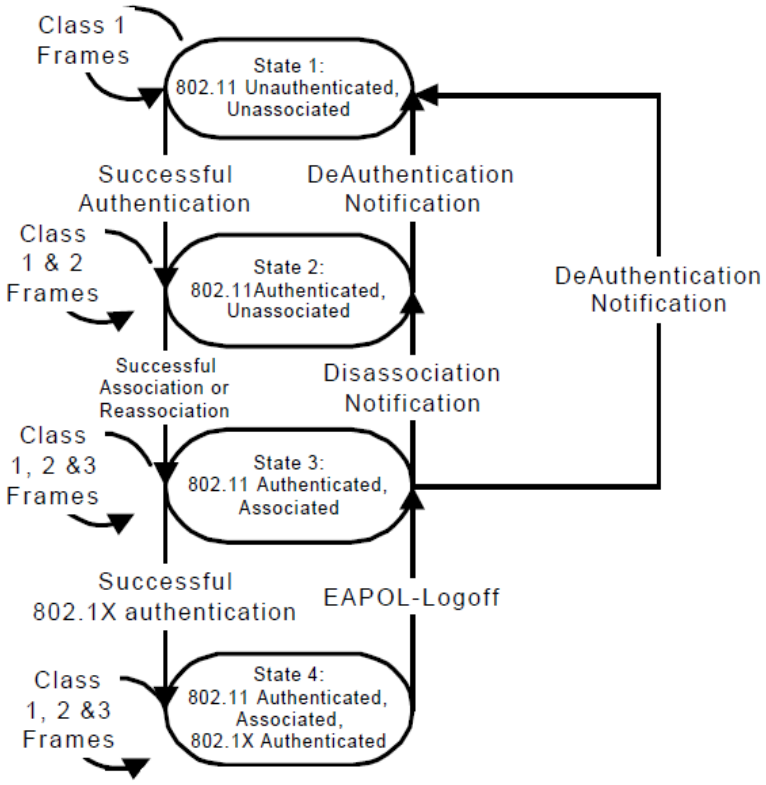

Figure 10: IEEE802.11 State Machine and State Transitions

All STAs who want to access the wireless network need to undergo the state transition sequentially from State 1 to State 4 by performing Open System authentication, association and the 802.1X authentication, as shown in the figure. In State 4 (after a successful 802.1X authentication), an encrypted communication session is established between the AP and STA with the transient keys generated during the four-way handshake process. A STA must stay in State 4 in order to retain the communication session. When the STA or AP wants to terminate the session, the initiating entity can send a deauthentication or disassociation message to the other entity to disconnect itself (hence transiting back to Stage 1 or Stage 2 respectively). A STA in State 1 and State 2 cannot participate in data communications until it has completed the 802.1X authentication (i.e., transit to State 4) again.

State transitions within the 802.11i state machine are controlled by management frames, and maliciously manipulating them can cause the protection measures to be averted. One of the major flaws of the state transition is that upon receipt of a deauthentication or disassociation frame the state machine unconditionally transits back to Stage 1 and State 2 respectively. According to the 802.11i standard, deauthentication and disassociation frames are notification frames and cannot be rejected. Further, those frames are not protected by any key material, so message integrity is not guaranteed. Hence, it is relatively easy for an attacker, pretending to be a legitimate STA or AP by spoofing the source MAC address, to send spoofed deauthentication or disassociation frames to other nodes, resulting in service disruption at the victim nodes.

By exploiting the vulnerabilities of unprotected management frames, DoS attacks become particularly threatening to the WLAN security and are extremely difficult to defend against. The following sections will describe the most common DoS attacks of this form in the existing WLANs.

3.2 Management Frame Flooding Attacks

3.2.1 Deauthentication/Disassociation Flooding

As previously mentioned, sending a spoofed deauthentication or disassociation frame can terminate the communication session of victim STAs[4]. Typically, the STAs would re-associate to regain service until the attacker sends another spoofed frame. By continuously flooding spoofed frames, the attacker can keep the victim STAs out of the service indefinitely. The attacker can also pretend to be an AP and broadcasts deauthentication or disassociation frames to all the associated STAs, causing all the STAs in the BSS to be disconnected from the network.

Those flooding attacks are easy to mount because the attacker can stop the communications using only limited resources without requiring any special technical skills or equipment. The attacker even does not need to break the authentication protocol or to obtain the shared secret keys between the STAs and the AP. Because the deauthentication/disassociation flooding attack is straightforward to implement, there have already existed several attack tools available on the Internet. Some popular implementations of such tools are “AirJack” [25], and “void11” [26].

To defend against such attacks, some research works have proposed solutions to explicitly authenticate management frames using digital certificates or public-key cryptography techniques [27, 28]. However, those solutions may not be preferable for enterprise deployments because managing a PKI (public key infrastructure) is expensive. The cumulative impact of security overhead introduced by the extra crypto-computations could also consume substantial amount of network bandwidth, and thus

[4] For more details on such attacks refer to [23].

creating a serious hindrance to achieving satisfactory QoS. Further, none of these research works actually address the support and integration for fast seamless handoffs.

A different approach proposed in [23] introduced a mechanism that delays the response of deauthentication or disassociation requests by queuing the received requests for 5-10 seconds. This gives the AP or STA an opportunity to observe subsequent packets from the sending party: if a data packet arrives after a deauthentication or disassociation request is queued, it indicates that the queued request is spoofed and should be dropped because a legitimate node would not normally generate data packets after sending the deauthentication or disassociation request. However, the major drawback of this approach is that the handoff delay would be significantly increased, so real-time services cannot be supported. Further, delaying the response of disassociation requests could possibly delay the re-association process during roaming, thus leaving the old association established for an additional period of time. This could confuse the routing updates in the DS that is necessary to deliver packets to the new AP after a handoff.

3.2.2 Authentication/Association Flooding

An AP can easily become an obvious target of DoS attacks. By exhausting the resources (e.g., memory or CPU) of an AP, network availability could be disrupted and effectively hinders any communications within the BSS. As being a stateful request-response model, the 802.11 management frames, in particular the authentication and association frames, appear to be a well suited target of exploitation for such attacks.

When a legitimate AP receives a spoofed (i.e., with a faked source MAC address) authentication request, it reserves memory for the request and replies an authentication response to the faked MAC address. Since there does not exist a wireless client that has this MAC address, the AP will not receive an Acknowledge (ACK) for the transmitted authentication response. The AP then retransmits several authentication response frames. If an attack is continuously sending out spoofed authentication frames, the victim AP has to keep allocating memory and replying multiple response messages for each received authentication request. This will soon overload the wireless bandwidth

and consume substantial amount of AP's resources so that the AP will have little or no resource to serve legitimate clients, causing them to suffer from poor communications or loss of the wireless access completely.

A legitimate AP also needs to maintain a table, called Association Table, in which the AP keeps track of the status of each associated STAs. An attacker can also flood the victim AP with spoofed association requests to fill up its association table with faked associations. When the AP's association table gets full, it will no longer be able to accept further associations, thereby denying access to legitimate STAs.

To prevent those flooding attacks, the AP needs to be able to distinguish between valid requests and spoofed requests. Unfortunately, the current 802.11i (and 802.11) standard does not support any means for validating those requests. At present, the most commonly used countermeasure for those attacks is to limit the rate of association requests. When the number of allowed associations is reached within a pre-defined time period, the so-called "DoS protection capable" AP simply blocks all further requests for the remaining time period [29]. The major drawback of this approach is that it does not distinguish between spoofed requests and legitimate requests. As a result, an attacker can easily bring the AP to a blocking state by using high flooding rates, effectively causing a new form of DoS condition to the WLAN.

Several research papers have proposed different solution to mitigate this issue [8, 27, 30-34]. Although they are able to mitigate those DoS issues, none of them actually address roaming support. In fact, most of them would increase the handoff delay substantially because of the introduced extra security mechanism that involves pattern matching, traffic learning, timing analysis, or heavy crypto-computations. Thus, practical deployment of those solutions in an enterprise network is not likely, and a better DoS mitigation technique that can actually support secure fast roaming is still a major research issue to be addressed.

3.3 EAP DoS Attacks

There are several DoS attacks that exploit the unprotected EAP messages in 802.1X authentication. The EAP frames sent during the 802.1X process are not protected by the 802.11i standard because an RSNA has not yet been established at this point. By flooding spoofed EAP frames, possible EAP DoS attacks include but are not limited to the following [35, 36]:

- Spoofed EAPOL-Start flooding which exhausts the AP's resources and prevents the 802.1X authentication from succeeding.
- Spoofed EAPOL-logoff attack to force a legitimate STA out of service.
- Faked EAP-failure attacks, where an attacker spoofs EAP failure frames to disconnect an authenticating session.

EAPOL-Start Attack

The EAPOL-Start message is sent from the supplicant to the authenticator to start the authentication process with the authentication server. Upon reception of such a frame the AP responds back to the supplicant with an EAP-Identify-Request and also processes some internal resource allocation. This vulnerability can be exploited by an attacker to flood randomly spoofed EAPOL-Start frame to the Access point, forcing the AP to allocate more and more resource until the memory runs out.

EAPOL-Logoff Attack

The EAPOL-Logoff message is sent from the supplicant to the authenticator indicating that it wants to terminate the authenticated communication session. Because the EAPOL-Logoff frame is not authenticated, an attacker can spoof the frame with the MAC address of a legitimate client and flood to the AP, causing the legitimate client being blocked from connection to the network.

EAPOL-Failure Attack

Similar to the EAPOL-Logoff attack, an attacker repeatedly injects forged EAPOL-Failure frames to disconnect a legitimate client from an existing communications session.

3.4 Four-Way Handshake Vulnerability

In addition to those previously described security threats, there is another potential DoS vulnerability at the four-way handshake stage. The four-way handshake is an essential component of the RSNA establishment, as it confirms the possession of the shared PMK in the authenticator and the supplicant, and derives a fresh PTK to secure the subsequent data communication.

As previously described in Section 2.4.3, the Message 2, 3, and 4 are authenticated and integrity protected by the fresh PTK, but the first handshake message is not protected with any key material for its authenticity and integrity. Therefore, an attacker can forge this messages to cause inconsistent keys being produced between the AP and the client because the nonce value in the faked Message 1 is different from the one used in Message 3 from the legitimate AP [15, 36, 37]. This effectively prevents the legitimate client from completing the RSNA.

Further, after each reception of faked Message 1, the client has to store ANonce and computes PTK. If an attacker can manage to successfully achieve multiple injection of faked Message 1 to the client (e.g., by flooding), memory exhaustion can result due to storing a large amount of ANonce and PTK values. This is depicted in Figure 11 below. This attack is particular threatening because it can be mounted without much effort, and a successful attack will ruin all efforts made in the prior authentication process.

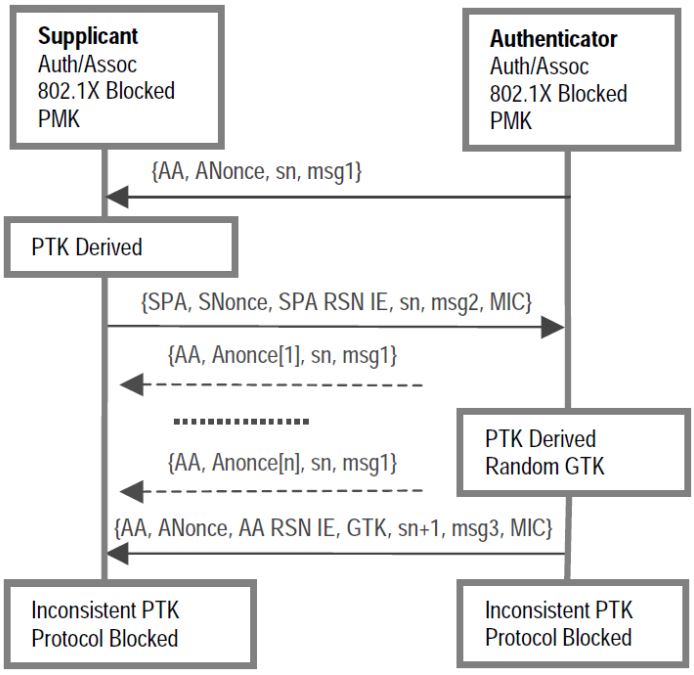

Figure 11: Message 1 flooding during four-way handshake

3.5 Chapter Summary

In summary, the current IEEE 802.11i security standard provides strong data encryption and mutual authentication, but it cannot prevent DoS attacks in WLANs exploiting the vulnerabilities of unprotected management frames. This chapter presents some of the most common DoS attacks against IEEE 802.11 WLANs. Unfortunately, those DoS attacks are threatening and easy to mount, and there already exists attack tools available on the Internet to perform most of those attacks. Therefore, it is necessary to have a security mechanism that can efficiently defend against those DoS attacks. For the solution to be useful in practical enterprise deployments, fast handoff and roaming support must also be considered. Although several solutions have been proposed by different research groups, none of them actually address the support for fast handoff and roaming related security issues. Hence, defending against DoS attacks in an enterprise environment still remains a challenge to be solved.

Chapter 4

DoS Attacks and Mitigation

Over the last few years, it has been a trend for companies and organisations to use IEEE 802.11 WLANs as a wireless extension to the core network for transferring data, which in many cases contain sensitive and critical information [38]. However, many organisations still overlook the potential impact of DoS attacks against their WLAN availability. Even with the latest WLAN security standard such as IEEE 802.11i, WLANs can still be vulnerable to DoS attacks and the results can be anything from degradation of throughput to a complete loss of wireless network access.

With the growing demand of WLANs in everyday life, there is a need to understand the implications of DoS attacks in these networks. Inspired by this fact, the thesis will study and examine the performance impact by DoS attacks in a WLAN. This chapter will first study the WLAN DoS attacks in more details and scrutinise those vulnerabilities. To quantify the performance affected by DoS attacks and obtain realistic results, a real-time testbed[5] is implemented to conduct empirical analysis to quantify DoS attacks based on measurements. This chapter starts by looking at common DoS vulnerabilities in an 802.11i based WLAN. Section 4.2 will describe the testbed setup, attack model, and present the experimental results. Insights obtained from the experiment results and the investigation of mitigation requirements and techniques will be discussed towards the end of the chapter in Section 4.3.

---

[5] In the thesis, the real-time testbed means a prototype network that is physically deployed and is equipped with required hardware and software modules.

4.1 DoS Vulnerabilities in 802.11i

In a RSN, the actual authentication takes place after the association. This is because the Open System authentication does not provide any security and is only used for the purpose of backward compatibility. Thus, clients are actually associated to the AP without any link-layer protection prior to the completion of an 802.1X authentication. The EAP authentication messages, which are encapsulated in 802.11 MAC data frames, are accepted by the AP via the 802.1X uncontrolled ports. The association exists only for a period of time sufficient for the 802.1X authentication to take place. If the 802.1X authentication does not complete within the time period, the client will be disassociated. However, this is in fact a DoS vulnerability because the AP's 802.1X ports cannot verify frame authenticity until the 802.1X authentication is completed. Moreover, the AP maintains considerable amount of state information after the association and before the completion of the 802.1X authentication. During this dark period, the AP needs to perform stateful operations (e.g., allocating memory) for requests that can not be validated. Hence, an attacker can flood an AP with spoofed authentication, association, deauthentication, or EAP requests to exhaust its resources from a single source with random MAC address.

The RSN performs the actual authentication using 802.1X/EAP and RADIUS via upper layer (above the IEEE 802.11 MAC sub-layer) authentication methods. This allows the network to authenticate clients based on credentials (e.g., username and password) rather than MAC addresses, which can be easily spoofed. Although this provides the flexibility of allowing authentication and key management functionality to be implemented without modifications to the existing IEEE 802.11 MAC sub-layer, this, however, opens up vulnerabilities for flooding attacks because there is no link-layer authentication and filtering capabilities. Moreover, the upper layer authentication exchange is stateful, so unprotected EAP frames are also vulnerable to flooding attacks.

From the above points, it suggests that the lack of per-packet authenticity and integrity in IEEE 802.11 management frames has been a key factor in many of the security problems. Those flooding-based DoS attacks described in the previous chapter

primarily exploit the lack of authenticity in management frames. Therefore, authenticity and integrity of management frames needs to be protected and assured in order to mitigate those DoS vulnerabilities.

To prevent DoS attacks on WLANs, the IEEE 802.11w working group is proposing cryptographic-based protection mechanisms to secure management frames. Those mechanisms are very similar to the protection of data frames using the keys derived for TKIP or CCMP. However, there are some limitations in the current version of 802.11w. First, all management frames sent or received by a station before keys are derived are unprotected. Secondly, 802.11w does not mention how to protect authentication request frame and association request frame, and it does not address how to prevent flooding attacks. Further, research works such as [39] have demonstrated that the 802.11w standard is effective only for low rate deauthentication and disassociation DoS attacks. Large-volume flooding with management frames against an AP can still lead to a DoS condition due to the heavy workload and overhead for computing the cryptographic key materials required to authenticate the spoofed frames.

4.2 Performance Impact of DoS Attacks

4.2.1 Testbed Implementation

To quantify the impact of DoS attacks on the throughput of wireless networks, a WLAN testbed is constructed to provide an experimental framework for measuring the effects of DoS attacks. As Linux open source community is growing rapidly, the fully functional 802.11i testbed built for this research use the Linux operating system and open source wireless device drivers and tools. The testbed is configured to use EAP-TLS as the upper layer authentication method and CCMP for data encryption.

The experimental setup consists of multiple APs connecting to the same subnet with the same ESSID, two wireless clients (supplicant), a wired traffic generator station (for generating network traffic), a RADIUS authentication server, and a wireless monitor

station that captures the wireless frames transmitted in the network. Figure 12 depicts the basic structure of the testbed.

All the stations are running on Pentium 4 (2.26 GHz CPU) machines with 512 MB RAM. Linux kernel 2.6.25 is installed on those machines. NTP (Network Time Protocol) is run on all stations to synchronize their clocks in order to allow accurate measuring and analysis of traffic latency and delay.

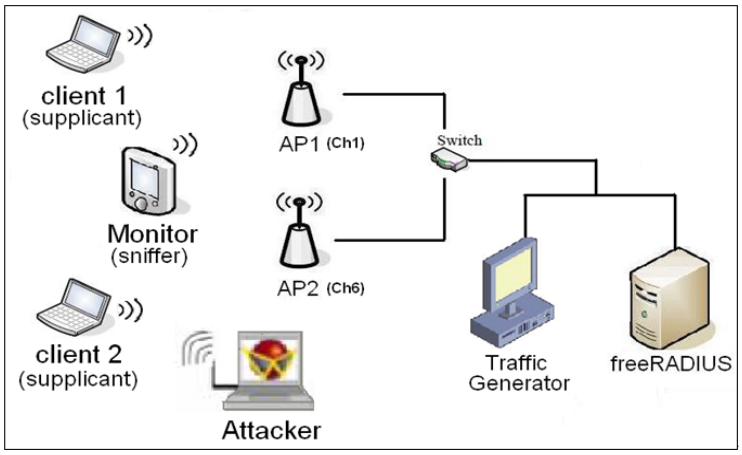

Figure 12: Testbed Structure

The RADIUS server, which provides authentication and authorization, is implemented using the open source tool - freeRADIUS[6]. Traffic analyser, Wireshark, is installed on the monitor station to analyse the captured frames. All the wireless stations, including the monitor, attacker and the two APs, are equipped with a D-Link DWL-G520 PCI wireless adapter card, which utilises Atheros AR5002G chipsets to provide IEEE 802.11g (2.4GHz) wireless access and hardware encryption functions (e.g., AES) for IEEE 802.11i security.

To configure Atheros chipsets to work on Linux systems, the wireless extension module in the Linux kernel is enabled, and an open source WLAN driver called

[6] freeRADIUS is available at: http://freeradius.org/

Madwifi Driver[7] (Version 0.9.4) is installed on all the wireless stations. This driver is later modified to implement the proposed APN authentication and fast handoff scheme which will be presented in Chapter 5 and Chapter 6 respectively. In order to get Madwifi driver working properly on Linux systems, an additional API package called the Wireless Tools for Linux[8] is also required to be installed on the stations. This tool allows the WLAN device parameters to be set on the fly in user space.

On the AP stations, an additional user space daemon for IEEE 802.11 access point management called hostapd[9] is installed. This open source implementation provides the 802.11 AP protocol module as well as IEEE 802.1x authenticator and RADIUS authentication client. It is configured to use IEEE 802.1X authentication and CCMP key management. To provide integrity, EAP-TLS upper-layer authentication protocol is configured and utilised.

The software architecture of the AP packages is shown in Figure 13. Inside the Linux kernel, WLAN driver (Madwifi) controls the WLAN hardware. User space applications such as Wireshark, TTCP, and Iperf use the raw socket interface provided by the kernel protocol stack to exchange frames with the driver. The user space wireless tool package uses the IOCTL interface to set or retrieve configuration information. The kernel wireless extension interprets the commands between the wireless tool package and WLAN device driver.

WPA supplicant[10], a user space supplicant implementation with support for WPA2 (IEEE 802.11i and RSN), is installed and configured on the supplicant station. This acts as the backend component controlling the wireless connection, key negotiation

[7] MadWifi is available at: http://madwifi-project.org/
[8] More information about Wireless Tools for Linux is available at: http://www.hpl.hp.com/personal/Jean_Tourrilhes/Linux/Tools.html
[9] More information on hostapd is available at: http://linuxwireless.org/en/users/Documentation/hostapd
[10] WPA Supplicant is available at: http://hostap.epitest.fi/wpa_supplicant/

with a WPA Authenticator (i.e., hostapd), and controls the roaming and IEEE 802.11 authentication/association of the WLAN driver.

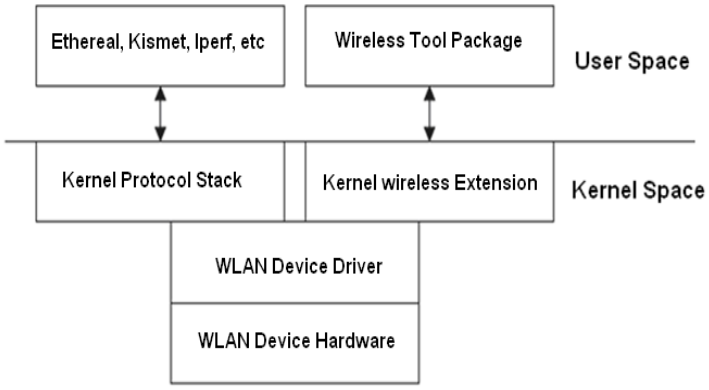

Figure 13: The architecture of the AP software packages

On the attacker station, attacking tools such as aircrack-ng and MDK3[11] are used to launch DoS attacks. Traffic generator and benchmarking tools such as Iperf and TTCP[12] are used for generating TCP/UDP traffic streams and measuring throughput. The network performance is analysed by sending traffic stream from the traffic generator station to the supplicants and taking the obtained throughput measurements using TTCP.

4.2.2 Experimental Results and Analysis

This section experimentally analyses the effect of WLAN DoS attacks, including deauthentication flooding, authentication flooding, and association flooding and briefly presents the experiment results. The open source tool MDK3 is used to launch those attacks on the attacker station. Figure 14 shows the different flooding mode available in MDK3.

[11] MDK3 is available at: http://homepages.tu-darmstadt.de/~p_larbig/wlan/

[12] More information is available at: http://www.pcausa.com/Utilities/pcattcp.htm

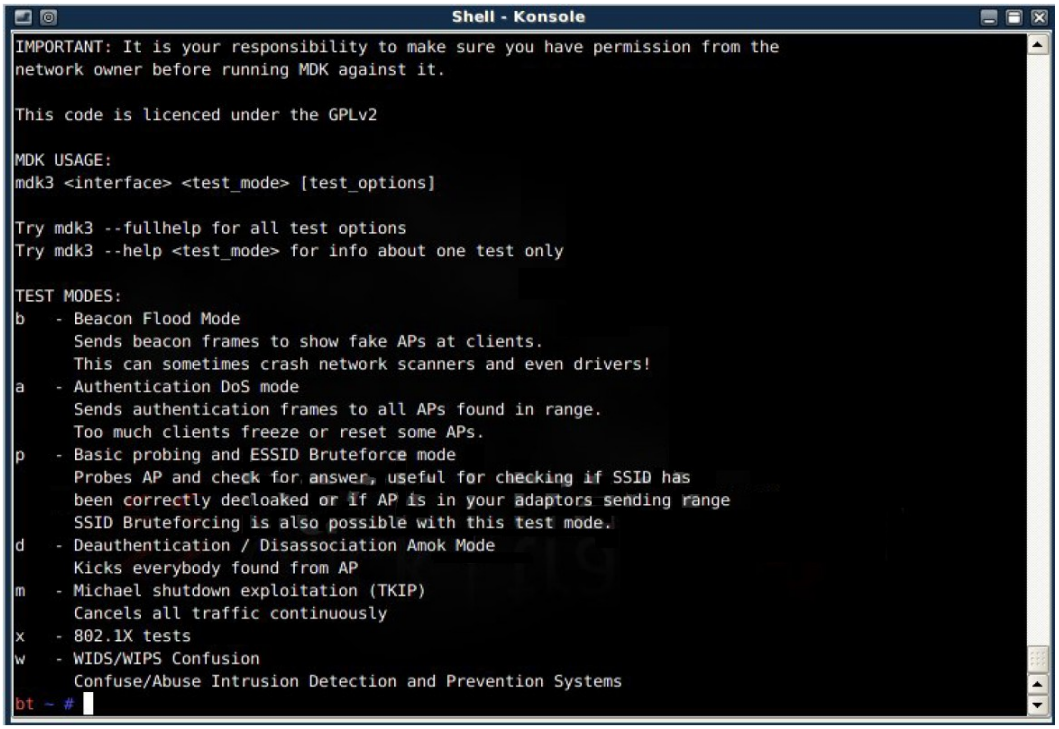

Figure 14: Attack options provided by MDK3

In the first experiment, deauthentication frames, which are spoofed with the real AP's MAC address, are continuously sent to the legitimate client from the attacker (fake AP) as depicted in Figure 15 below.

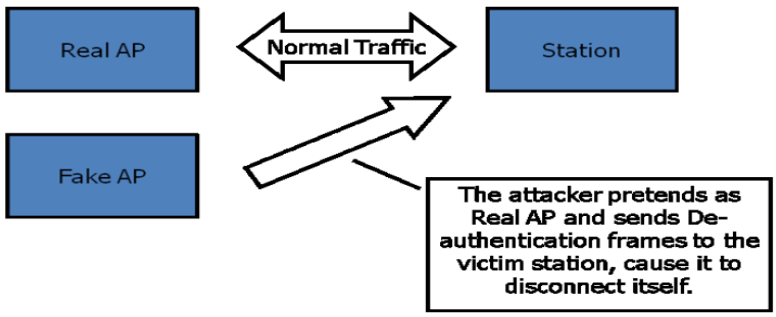

Figure 15: Deauthentication Flooding Attack Model

The following diagram shows the deauthentication flooding in action with MDK3. The left hand column shows the client's MAC address and the right hand column shows the APs MAC addresses. All the clients are being forced to disconnect from their original APs until the attack is manually stopped.

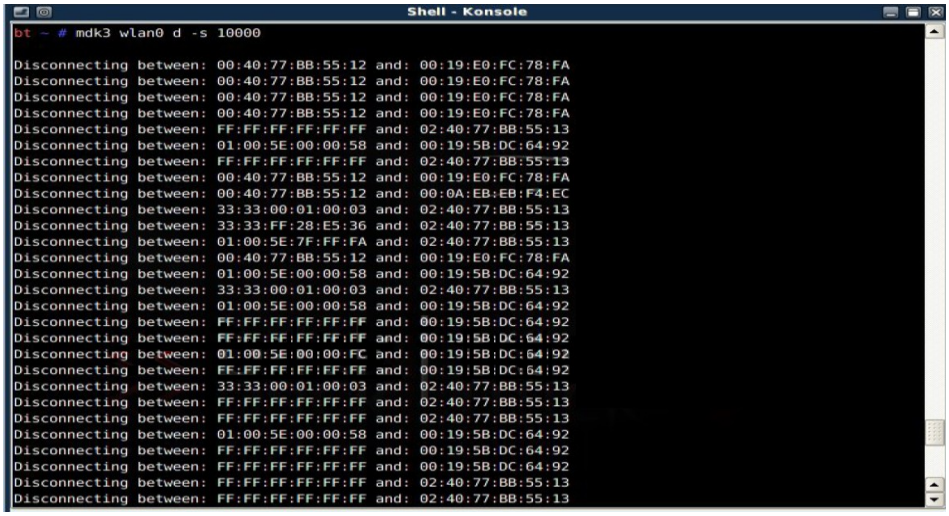

Figure 16: Deauthentication/Disassociation Attack Mode

The experiment even demonstrates that flooding at a low rate of one frame per second is well enough to block the victim client from connecting to the AP. When the client attempts to reconnect, another spoofed deauthentication frame from the attacker will immediately end the new connection. As a result, the throughput of the client drops to zero over the duration of the attack.

The MDK3 tool can be run in the Authentication DoS mode which generates spoofed authentication requests and continuously sends them to a target AP. Under this attack, the AP becomes too busy processing the fake requests to provide normal service to legitimate clients. Figure 17 shows the options available in the MDK3 Authentication DoS Mode.

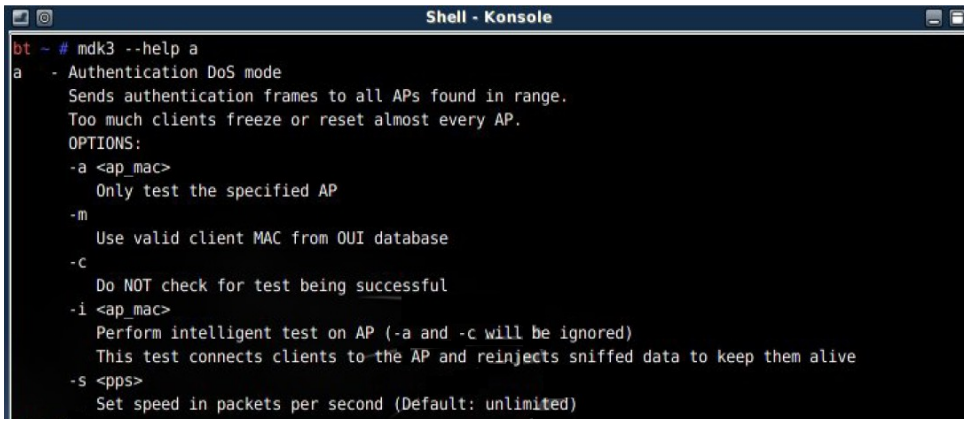

Figure 17: Authentication DoS Mode Options

By default, authentication requests are sent to the specified AP (the MAC address given with the –a option in the command line) at its possible maximum rate, and the status of the AP is reported after every 500 packets are sent. When the attack is first launched, the AP can still respond to new legitimate connections. After running for 5 minutes, 38000 connected clients caused the AP to freeze. A reboot on the AP station was then required to bring the AP back to the normal state even though the attack was stopped already. Figure 18 shows the authentication DoS attack in action.

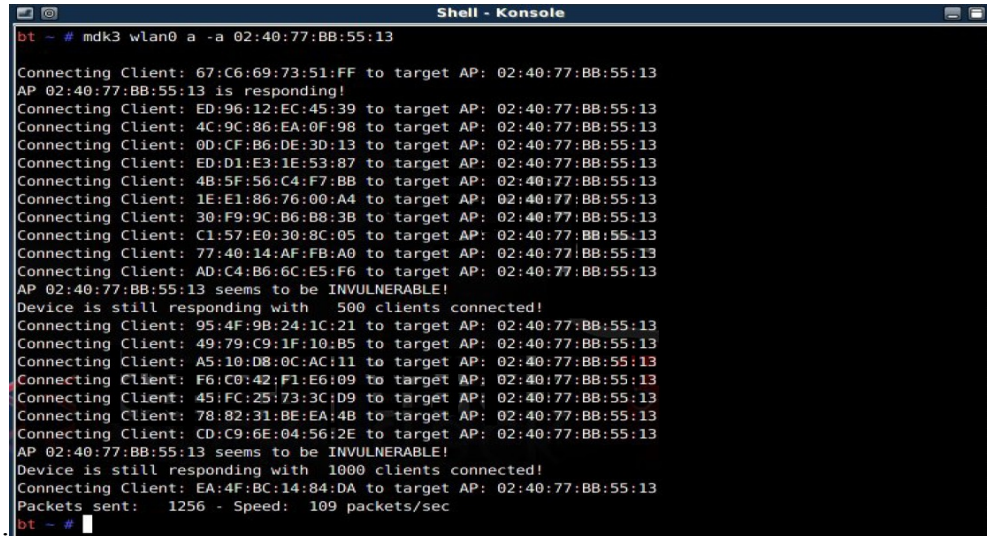

Figure 18: Authentication DoS in action

For throughput measurements, a continuous stream of TCP packets is generated from the traffic generator STA to the mobile clients. When the client receives the spoofed deauthentication frame, it immediately disconnects and goes to the unauthenticated state. Figure 19 (a) shows that the client's throughput remains zero for the duration of the attack. The throughput goes back to normal after the attack is stopped and the client reconnects to the AP.

Similar result is obtained for authentication flooding. As Figure 19 (b) shows, during the period of authentication flooding (from 10th second to 25th second), there are only a little legitimate packet transmissions against the huge amount of flooding packets. This is due to the effect that the flooding traffic consumes most of the resources of the AP and the AP is completely busy responding to the spoofed frames. The AP's throughput drops to less than 10% of the normal capacity during these 15 seconds, and the

throughput remains low for a few seconds after the attack is stopped. If the duration of the attack is long enough, the AP will eventually freeze and require a reboot. Flooding with association requests using the attack tool, void11, gives a very similar result to Figure 19 (b).

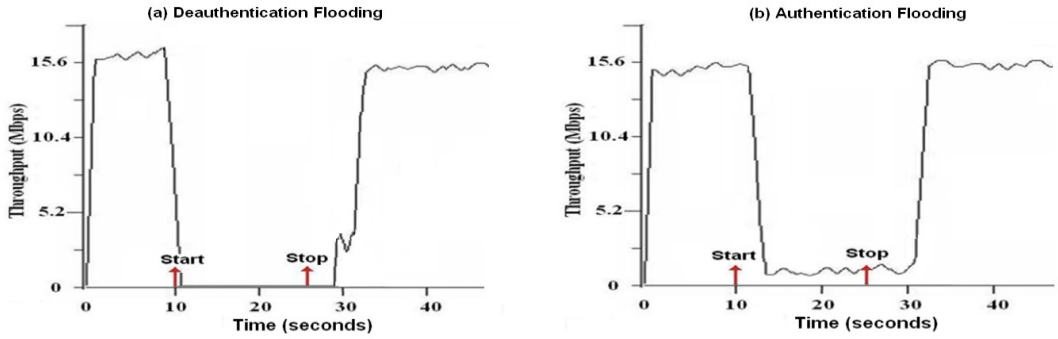

Figure 19: Throughput measurements for (a) deauthentication flooding and (b) authentication flooding attacks

In summary, the result suggests that deauthentication flooding is the most effective attack against WLAN throughput than authentication flooding and association flooding. Flooding with spoofed deauthentication frames can completely bring down the wireless network and no legitimate users can continue using the network at all. Those experiments demonstrated that it is easily to launch DoS attacks on the WLAN, and without any further DoS protection, the WLAN is very vulnerable to flooding attacks.

4.2.3 Frame Protection with IEEE 802.11W

To prevent those DoS attacks, the AP requires the capability to distinguish between legitimate requests and spoofed requests. The IEEE 802.11w standard is designed to achieve this and further enhance MAC layer security by providing authentication and encryption protection to some of the 802.11 management frames.

According to the standard, an 802.11w compliant STA should be able to distinguish whether a management frame (e.g., a deauthentication request) comes from the AP it is connected to (if the AP also supports 802.11w) or it is spoofed by the attacker masquerading as the AP. The STA should ignore the spoofed deauthentication frame to prevent the DoS attack from being successful.

An experiment was conducted to verify the effectiveness of 802.11w protection against those flooding DoS attacks. Both the wpa_supplicant and the hostapd are configured to utilise the 802.11w management frame protection. With the 802.11w protection enabled, the same tests performed previously were repeated and the output of the hostapd log indicates that the STA was able to identify and deny spoofed frames. However, this process involves authenticating the frames cryptographically (using hashing, encryption, and decryption) and encapsulation/decapsulation, so the verification actually consumed significant resources of the AP. As the flooding rate increases, throughput degradation becomes more and more apparent. Figure 20 summarises the TCP throughput performance of the AP, which is under deauthentication flooding using different flooding rates, ranging from no flooding (normal throughput) to up to 400 fps (frames per second).

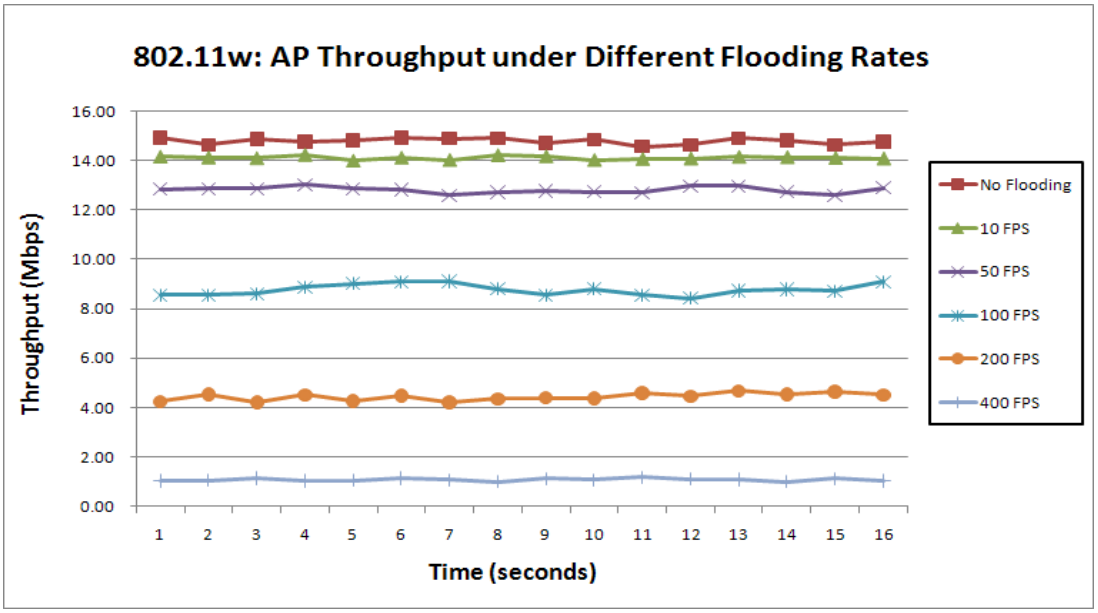

Figure 20: TCP throughput degradation under various flooding rates against 802.11w protected AP

When the flooding rate is at 10fps, little performance impact was noticed. As the rate increases to 400 fps, the throughput drops more than 90% from the normal throughput when there is no flooding traffic.

In summary, the result proved that the 802.11w frame protection is able to successfully identify and drop spoofed frames. However, high rate flooding can still exhaust the

AP due to enormous crypto processing overhead involved so that it has little or no resources to continue processing regular data communications. As a result, this becomes another form of DoS condition.

4.3 Mitigation Requirements

Several research work has proposed solutions to mitigate DoS attacks by authenticating management frames with different cryptographic techniques [40-44]. However, there are several limitations with those solutions as they only focused on improving security with additional cryptographic operations, whereas lacking detailed quantification of the associated security overhead for meeting the required QoS for real-time applications. Moreover, those solutions may not be effective against high rate flooding attacks, and they may not be capable of providing satisfactory handoff performance.

In order to enhance security with DoS mitigation capabilities, a link-layer protection against spoofing and eavesdropping is needed. Ideally, this protection should take place as early as possible before the AP allocates resources for establishing any connection state. DoS attacks are likely to succeed in WLANs mainly because they can be mounted prior to the completion of an 802.1X authentication. If a lightweight, link-layer authentication mechanism is possible to take place before the association phase, none of the aforementioned attacks can occur. However, providing link-layer protection in WLANs is a challenging task.

In contrast to wired networks where executing expensive computations on end devices normally does not pose performance issues due to comparable hardware capabilities, utilising cryptographic-based security measures in wireless networks can lead to various performance related problems. The flooding experiment performed in Section 4.2.3 (Frame Protection with IEEE 802.11w) is a good example of this. Protection measures such as the 802.11w that involve stateful protocol execution, high crypto-overhead, or complex message exchanges are likely to introduce new DoS vulnerability. The computation resource and the channel utilisation overhead of the mitigation solution should be kept minimal because the clients are usually resource-

constrained devices. Further, millions of WLAN products have been released to market and are in use. Any new solution should be compatible with the existing 802.11 standards and require minimal modifications.

Unfortunately, most of the WLAN service providers or wireless Internet services providers have abandoned link-layer security, and instead, they enhance security by using proprietary solutions based on Web-based authentication. This trading of link-layer security can have a high impact on users' security and introduce new vulnerabilities as discussed in [45].

To properly mitigate link-layer DoS vulnerabilities and allow interoperability between existing WLAN devices, any new mitigation solution should be based on the following three identified requirements:

1. **Provide frame authentication with a simple and lightweight authentication mechanism:**

   This will ensure management frames are authenticated in order to prevent MAC spoofing based flooding attacks. Considering high rate flooding can degrade the overall network performance if excessive resources are used to validate the frames, the solution should be lightweight so frames can be processed quickly without affecting too much of the network performance.

2. **Use stateless protocol execution in order to be DoS resilient.**

   To prevent the solution from being a new DoS vulnerability or performance bottleneck, stateless operations should be used to provide initial frame authenticity verification before stateful operations are performed to process further authentication requests and handle stateful information.

3. **No modifications to the current 802.11 state machine.**

   For practical deployment considerations, the solution should not change the existing 802.11 protocol state and needs be compatible with legacy devices. The solution could be implemented only by upgrading device firmware or applying a patch to the driver.

4.4 Chapter Summary

WLANs are more susceptible to DoS attacks than wired networks because the wireless medium is not confined to physical boundaries such as wires and buildings. The authentication/encryption of management frames was never a part of the original 802.11 specification either. This makes it easy for malicious users to perform DoS attacks by spoofing and/or flooding with management frames. Tools that exploit this vulnerability to launch DoS attacks are freely available on the Internet. Section 4.2.2 illustrated how easy it could be to successfully launch DoS attacks using those attack tools, and the obtained results showed that flooding attacks are very effective against WLANs that do not have any additional link-layer DoS protection.

Although later in 2009 the IEEE 802.11w amendment was introduced to mitigate certain types of WLAN DoS attacks, there are still several shortcomings. First, 802.11w applies the same cryptographic protection for data frames to management frames, so only those management frames such as deauthentication and disassociation frames that are transmitted after the completion of an 802.11i RSN key derivation can be protected. Therefore, authentication and association flooding attacks are still possible in a RSN with the 802.11w enabled. Second, the 802.11w does not perform well under high rate flooding attacks, as indicated in Figure 20. Therefore, an effective DoS vulnerability mitigation solution still needs to be investigated. Based on the requirements identified in Section 4.3, the research will design and implement a novel lightweight authentication scheme, which allows frame authentication with stateless operations before the AP allocates resources for establishing a connection. The design and detailed description of the proposed authentication solution will be presented in Chapter 5. The implementation and related experimental results will be presented in Chapter 8.

Chapter 5

DoS Mitigation with Client Puzzles

To mitigate resource depletion DoS attacks, the ideal solution is to have an efficient and effective filtering mechanism that can distinguish between legitimate traffic and flooding traffic [46]. However, the major obstacle to defending against DoS attacks, particularly in WLANs, is the lack of this ability to identify if a frame is from an attacker or from a legitimate user.

In wired IP networks, traffic filtering can be achieved using authentication schemes based on the concept of "predictive nonces" [47]. An example of this DoS mitigation technique is this thesis author's previous published work on the mitigation of SIP (Session Initiation Protocol) flooding attacks in VoIP systems [48]. The paper proposed a stateless firewall nonce checking mechanism as an extension to the existing (stateful) SIP digest authentication. The basic idea is that when a server receives a request, it challenges the client with a nonce that is computed as a function of those headers in the request (e.g., source IP address) which it knows to be invariant when the client resubmits its request; when the new request with the nonce attached arrives, the server re-computes the nonce using the same set of headers in the same way. If the headers have not changed and the request comes from the same original source, the resulting nonce will be identical to the one attached in the request. By checking the validity of the nonce the server can distinguish spoofed requests or legitimate requests. The security of this technique is based on the fact that even if an attacker could determine the algorithm for computing a nonce, and therefore determined what nonce to use for a message it modified, but because of the spoofed source address, the attacker would not have access to the digest response from the server which was computed using that nonce.

Unfortunately in a wireless network, this technique does not work due to the broadcast nature of message transmissions. An attacker may be able to obtain all the transmitted information by sniffing wirelessly within the range without regard whether the attacker's MAC address is spoofed. To mitigate DoS attacks in WLANs, a different approach is needed which takes into account the broadcast nature of information transmission. Based on the concept of predictive nonce and the requirements identified in Section 4.3, a novel link-layer authentication scheme based on client puzzles is proposed. The proposed authentication scheme, which will be referred to as APN (Access Point Nonce) authentication, replaces the Open System authentication exchange used in the initial RSNA establishment, and provides a means for identifying legitimate clients without storing any state information on the AP.

The APN authentication assumes that an initial pre-shared secret is established between each client and the APs in the corporate network. The distribution of the shared secret is not the focus of the research, and it is assumed that the shared secret could be securely configured by using device management tools that are commonly used in enterprise networks. Unlike WEP or WPA pre-shared keys, which will not be changed, the shared secret used in the APN authentication can be dynamically updated every time a RSNA is established successfully. The shared secret only needs to be configured once when a client device first connects the WLAN, so there are no security concerns, such as key leakage or manual key management, like most of the shared key authentication schemes (e.g., WPA-Personal) would typically have.

In the proposed APN authentication, the AP issues each client requesting network access a unique "client puzzle". A client puzzle, formulated using a secret only known to the AP, timestamp, and an AP generated predictive nonce (i.e., APN) computed using some client specific information, is a solvable cryptographic problem a client must solve in order to have the AP or network server resource allocated for its connection. A legitimate client who knows the pre-shared secret is able to immediately obtain the puzzle solution without spending extra resources solving it, whereas an attack would have to spend a huge amount of time solving a puzzle for each request frames transmitted in order to launch a flooding attack. The APN authentication also

provides a means to allow the client and the AP to authenticate subsequent layer 2 frames, including management frames and EAP messages, with a technique that uses the hard factorisation principle that will be described in Section 5.1.3.

The proposed APN authentication provides a lightweight, effective link-layer authentication mechanism. By incorporating the benefits of predictive nonce, client puzzle, and hard factorisation, the solution is capable of preventing various forms of resource exhaustion DoS attacks, as well as mitigating problems that may arise from the link-layer vulnerabilities of unprotected management frames.

In this chapter, Section 5.1 first presents some related work in client puzzles and describes the puzzle construction and verification process, as well as the hard factorisation problem, which is commonly used in cryptographic techniques such as RSA [49]. The proposed APN authentication will be described in details in Section 5.2. The security of the APN authentication scheme will be analysed and discussed in Section 5.3. Finally, Section 5.4 concludes the chapter with a summary. Further implementation details and experimental results will be presented later in Chapter 8.

5.1 Related Work

One technique that has been used in wired networks to address resource depletion DoS attacks is based on the principle that a client, requesting for network service, should commit a certain amount of its resources first before the server does. A common implementation of this concept is a client puzzle. A puzzle presents a cryptographic problem that the client needs to solve in order to show its legitimate interest of obtaining network services. Juels and Brainard in [50] are the first to introduce client puzzles to prevent connection depletion attacks in wired networks. Over the past few years, various client puzzle schemes have been proposed [51-55], but the basic idea is the same: when a server is under attack, it sends out a cryptographic puzzle for the client to solve before allocating resources for that client.

A puzzle is created by taking a difficult problem from an appropriate cryptosystem and making it "feasible" to solve by providing extra information that will assist the client in finding the solution. The server can also adjust the difficulty of the puzzle depending on its traffic condition. The client must solve the puzzle within a certain time interval. The client who presents a correct solution to the puzzle will be granted network access, and the server starts to commit its resources serving the client.

For an attacker to launch a DoS attack against the server, a large number of puzzles need to be solved within a short period of time which is practically infeasible with limited computational resources. In contrast, a legitimate client only needs to solve one puzzle for its session. This method is effective in separating flooding requests from the legitimate traffic [54].

Client puzzles have been proposed to defend against resource depletion DoS attacks in the context of TCP [50, 54, 55], authentication protocols (such as [56]), and TLS [57], etc. Although using client puzzles to mitigate DoS attacks is not a new concept, incorporating client puzzles in WLAN link-layer authentication is a relatively new attempt.

5.1.1 Client Puzzle Characteristics

The main idea of client puzzles is to allow the server to generate puzzles that the client must solve before the server creates any protocol state or performs stateful operations. To prevent DoS stacks, puzzles should have the following characteristics [50, 51]:

1. **The puzzles should not introduce any new DoS vulnerability.** Generating a puzzle and verifying a puzzle solution should be inexpensive for the server so that receiving a flood of requests will not exhaust its resources. The process needs to be stateless so the server does not need to store any information about the client.
2. **The difficulty of the puzzle should be easy to adjust.** In order to provide graceful degradation of services, the server should be able to increase and decrease the computational resources required from the client in solving a puzzle as the server's load increases and decreases.

3. **The puzzles should require the client to commit adequate computational resources.** First, it should not be possible for an attacker to keep a table of known puzzles and pre-compute solutions. Secondly, the puzzles should be time-dependent so that the client has a limited amount of time to find the solution.

Most of the client puzzles that have been proposed require a puzzle to be solved by reversing a one-way hash function with brute force searching. The hash function used for client puzzles does not need to be collision resistant, but it is required to be resistant to inversion [58]. For example, fast hash functions like MD5 [59], or SHA [60] are the common candidates. Depending on the difficulty of the puzzle is set, solving a puzzle can be a trivial or infeasible task. In general, adjusting the difficulty of a puzzle means how much information about the solution is revealed to the client. The following section describes the typical process of computing a puzzle and solving a puzzle.

5.1.2 Puzzle Construction and Verification

The following notations are used to illustrate the client puzzle protocol.

X <i>	The i^{th} bit of a bitstring X
X <i, j>	The sequence of bits: X <i>, X <i+1>, …, X <j>
$\mathbf{E_k}$ **(*X*)**	Encrypt X with the key k
$\mathbf{D_k}$ **(*X*)**	Decrypt X with the key k
S	A secret bitstring which is only known to the server
h	A non-invertible hash function
l	Length of the digest output of h
||	Concatenation

Table 2: Notation for describing client puzzle protocol

A typical puzzle construction is as follows [61]:

First the server generates a bitstring x, which is computed as the hash of a set of service parameters and a server secret *s*. This hash digest is referred to as a pre-image.

The pre-image is then hashed again to produce a bitstring *y*, which is referred to as the puzzle image. The puzzle *p* consists of a portion *x* <*k*+1, *l*> of the pre-image, which is denoted as *x'*, along with the image *y*. The puzzle is depicted as the shaded bitstrings in Figure 21.

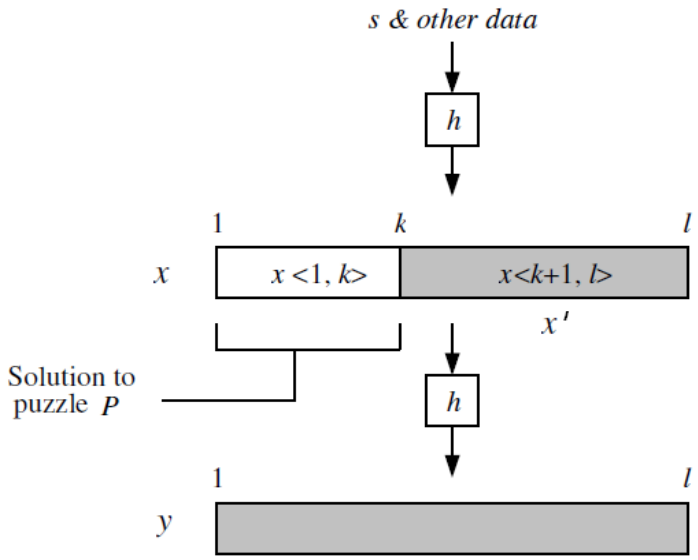

Figure 21: Construction of a client puzzle

To find the solution to *P* the client must find a value for *x* <1, *k*> (i.e., the missing portion of the pre-image) so that *y* = h (*x* <1, *k*> || *x* <*k*+1, *l*>). Thus, the difficulty of the puzzle is equivalent to the hardness of brute-force searching a space of all 2^k possible answers, where each step in the search requires a hash computation. The difficulty of the puzzles can be adjusted by changing the value of *k*.

When the solution is found, the client constructs the pre-image *x* (by concatenating the solution with the given partial pre-image *x'*) and sends it along with the given *y* to the server. The server also computes a pre-image for the client and verify the puzzle solution by checking whether *x* is a valid pre-image of *y* (i.e., checking h(*x*) = *y*). Those two easy hash operations have almost no computational cost to the server in verifying a solution. Using client puzzles, legitimate clients will experience only a small degradation in connection time and attackers must consume a large amount of resources in order to create an interruption to the network availability.

5.1.3 Hard Factorisation

The proposed APN authentication requires another useful technique, which is a well-known and commonly used mathematical hard problem in cryptosystems called hard factorisation, in order to provide the capability of per-frame authentication.

Hard factorisation is a prime decomposition process which requires splitting a large integer into factors that are prime numbers. By the fundamental theorem of arithmetic, every positive integer has a unique prime factorization. Multiplying two prime integers together is easy, but factoring the product of two (or more) prime numbers is much more difficult. The difficulty of factoring large prime numbers is the underlying building block upon which several public-key cryptosystems such as the RSA algorithm [49] use. Thus, when given a large number $N = P \times Q$ where P and Q are large prime numbers, it is computationally infeasible to find P and Q. On the other hand, when given P and Q it is easy to compute N.

In the context of the proposed APN authentication, the product N is referred to as an "identity token", and the prime numbers P and Q are referred to as "validating keys", for which can be used to validate an identity token. The validating process is simply a division operation, which is a trivial task for a computer. If $P \mid N$ (P divides N), then the validating key P proves the validity of the identity token N. More details on how the identity token and validating keys are used in providing per-frame authentication will be presented in Section 5.2.2.

5.2 Proposed Authentication Scheme

In WLANs, the authenticity of the client requesting services is uncertain, so the server could compromise its resources processing spoofed requests, ending up depleting all the resources. Thus, it is important for any DoS mitigation solution to avoid allocating resources before the client is proven to be legitimate.

The best solution to eliminate those vulnerabilities is to provide a means for the network to perform stateless operations first to verify the authenticity of the received request frames, before processing the requests with stateful operations. If the authenticity verification fails, the frame should be dropped immediately without having to perform any (stateful) upper-layer authentication in a RSNA.

To achieve this goal, a novel link-layer authentication scheme called Access Point Nonce (APN) authentication is proposed to replace the existing Open System authentication that a RSNA establishment uses. The APN authentication scheme takes the lightweight and stateless properties from the previously proposed firewall nonce authentication for SIP proxy, and extends its functionality into providing per-frame authentication by utilising the hard factorisation technique. The identity token of a connecting client and the AP are exchanged as a part of the initial authentication phase. To make it practically infeasible for an attacker to perform a successful flooding based DoS attack on the APN authentication exchange, client puzzles are used to protect the identity token exchange so that the AP only stores the identity token of a client, along with the client's information, after the client has successfully solved a puzzle. The puzzle construction is based on the method proposed by Juels and Brainard [50], with slight modifications to incorporate a predictive nonce generated by the AP (i.e., APN).

The following sections will first describe how APN authentication works to provide stateless authentication, and then show how it is extended to support per-frame authentication using the exchanged identity tokens after a successful APN authentication exchange.

5.2.1 APN Authentication Procedure

The proposed APN authentication scheme assumes that a temporary shared secret (or shared key) exists between a client and an AP. The purpose of this shared key is to provide an initial trust relationship between a client and the AP with which identity tokens will be exchanged. The shared key is only used for a short period of time, as it will be dynamically updated on both entities after a successful 802.1X mutual

authentication. Thus, the shared key does not need to withstand active attacks, and should be simple and efficient to implement.

The proposed APN authentication procedure is depicted in Figure 22 below. The client (or supplicant *s*) initiates the APN authentication after the discovery phase in which the target AP's MAC address has been learnt from the received probe response. The client first generates two large primes (P_s and Q_s) and computes its identity token N_s ($N_S = P_S \times Q_S$), and then it sends an initial authentication request to the target AP. This request frame encapsulates the client's MAC address and its identity token *Ns* as shown in the figure as message 1.

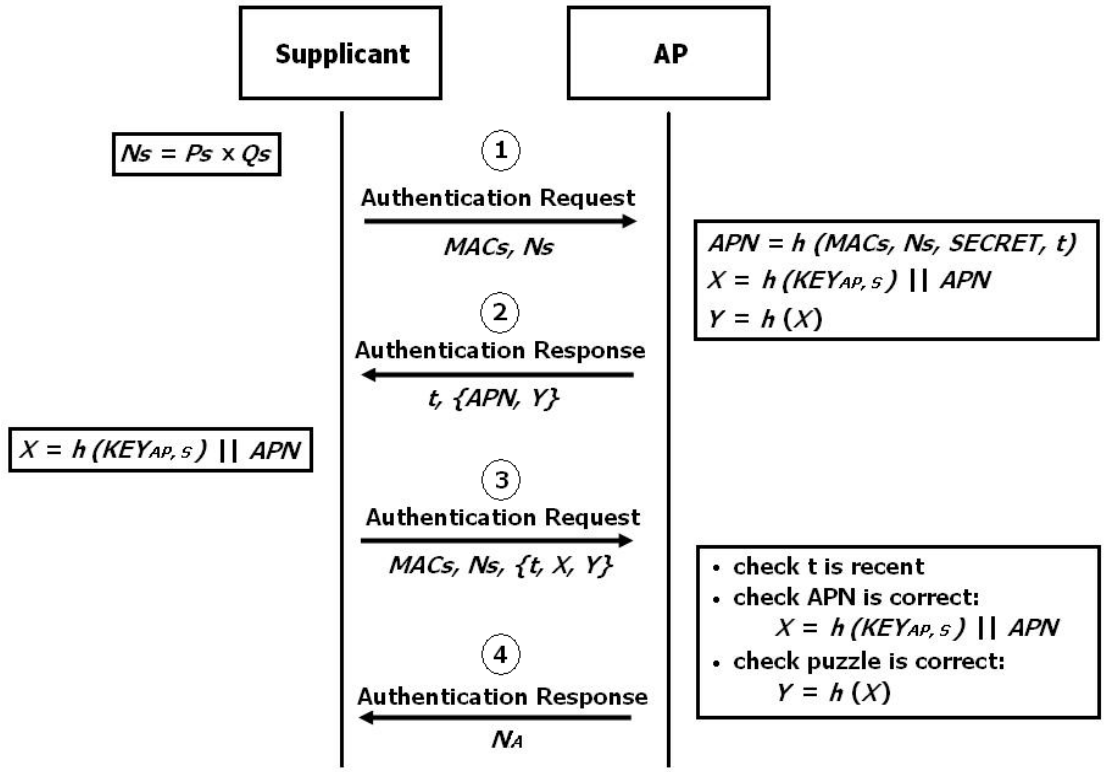

Figure 22: The proposed APN authentication procedure

Upon receipt of an authentication request, the AP checks whether there is an AP nonce (APN) attached to this request. If the received request does not include an APN, the AP generates one, which is the result of a cryptographic hash function computed over the client's MAC address (MACs), identity token (*Ns*) , a secret that is only known to the AP, and the current timestamp. The secret prevents nonce forgery by a third party,

because it is virtually impossible to generate a nonce that can be accepted by the AP without knowing the value of this secret. Including the timestamp prevents replay attacks with a sniffed or used APN. The APN generation is shown below:

$$\mathbf{APN = h\ (MACs, Ns, SECRET, t)} \quad \ldots\ldots\ldots\ldots\ldots\ (1)$$

To provide a binding between the APN and the identity of the client without trusting its MAC address, which can easily be spoofed, the AP validates the client by challenging it with a puzzle constructed in a way that a legitimate client who knows the shared key (denoted as $KEY_{AP,S}$) can easily solve it. To do this, the AP first generates a pre-image by hashing the client's shared key and concatenating the output with the APN, then a puzzle image *Y* is computed by hashing the pre-image:

$$\mathbf{Pre\text{-}image:}\ X = \mathbf{h\ (KEY_{AP,S})\ ||\ APN} \quad \ldots\ldots\ldots\ldots\ldots\ldots(2)$$

$$\mathbf{Image:}\ Y = \mathbf{h\ (}X\mathbf{)} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(3)$$

To set the puzzle difficulty, the first 128 bits (i.e., the key digest part) of *X* is removed, which effectively means that the APN itself is used as the partial pre-image (*x*') that forms the puzzle together with *Y*. The client puzzle construction process is depicted in Figure 23, where the shaded bitstrings represent the puzzle.

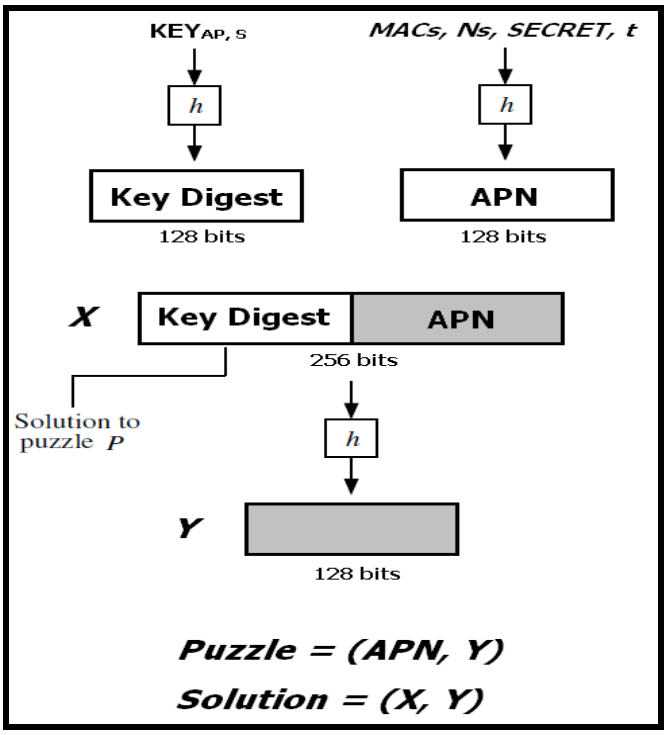

Figure 23: APN Client puzzle construction process

After constructing the puzzle, the AP attaches the timestamp and the puzzle in an authentication response frame and sends it to the client as depicted as message 2 in the figure. In the authentication response frame, the status code field is used to indicate the result of the previous authentication request. Since a puzzle challenge is required and there is no APN and puzzle solution attached in the initial request, the status code of the response is set to 27 (reserved code), indicating that an APN puzzle challenge is required. The AP then terminates the session without storing any information.

The client then obtains the puzzle from the received response frame and computes the puzzle solution by hashing the shared key. Because only the legitimate client knows the shared key, solving the puzzle involves nothing more than generating the key digest, meaning that only one hash operation is needed on the client station; whereas for attackers who do not know the shared key, solving the puzzle involves brute-force searching all the 128 bit key space, which is practically infeasible to do even with a modern processor. The shared key will be refreshed every time an 802.1X mutual authentication is completed. This update normally happens within less than 400-500ms after the key digest is first produced. If an attacker can somehow manage to reverse the 128-bit hash to obtain the key, it is unlikely that this could be achieved within such a short period of time.

After solving the puzzle, the client transmits another authentication request, depicted as message 3, which contains the same identity token and timestamp, together with the puzzle solution (X and Y). The following checks will be performed sequentially on the AP in a stateless fashion to ensure that the request is legitimate:

- **Timestamp:**

  The AP checks the attached timestamp t against the current time to make sure that the APN is recent enough. This prevents replay attacks and any reuse of puzzles. A request frame with an expired timestamp is immediately dropped.

- **APN:**

  The AP then re-computes an APN on the fly using the same computation method in Equation 1. If the computed APN matches the last 128 bits of X, the APN is

proved valid. An invalid APN could mean that the nonce is forged or the request has been modified. A frame with an invalid APN is immediately dropped.

APN provides a binding between the client's MAC address and its identity token. However, MAC addresses cannot be trusted in WLAN, so this binding needs another layer of protection to ensure the identity token is mapped to a legitimate identity. This is achieved by embedding the APN in a client puzzle so that it is bound to the trusted shared key of the client who can solve the puzzle.

- **Puzzle verification:**

  Having a correct puzzle solution indicates that the client is actually the one who sent the initial authentication request and can be considered legitimate. The puzzle binds the client's identity token with the trusted shared key. Thus, after a client successfully gives a puzzle solution, the attached identity token has been verified belonging to a legitimate client and can be used for subsequent frame authentication.

If all the three checks are successful, the client is considered legitimate and will be allowed to proceed with stateful operations such as the association and the upper layer authentication. The AP concludes the authentication by sending another authentication response (depicted as message 4 in the figure), which contains the AP's identity token N_A to the client. Upon receiving the response frame, the client stores the AP's identity token, which will be used to authenticate subsequent frames coming from the AP.

Now the APN authentication exchange is completed. The association phase then comes in and the AP stores the client's connection status and the identity token in the AID (Association ID) table, which is maintained by the AP for all the associated clients. The AP only needs to maintain one copy of its identity token N_A for all the associated clients, whereas each client computes a new identity token N_S every time it associates and re-associates to an AP.

In order for a client to use APN authentication, the AP indicates that it supports APN authentication in its beacon frames. This will allow the use of the original Open System authentication for backward compatibility.

5.2.2 Per-Frame Authentication with Validating Key

The APN authentication scheme not only validates the station's MAC address, but also provides a binding between its MAC address and its identity token. This allows a recipient to assure the source authenticity of received frames by examining the sender's identity token with a given validating key attached in the frame. A symmetric key cryptography (e.g., AES) is used to protect the identity token attached in the transmitted frames. Therefore, after the complete APN authentication exchange is successfully performed, subsequent unicast and broadcast (including multicast) management frames can be effectively authenticated.

For a frame to be authenticated, the sender encrypts one of its validating key (either *P* or *Q*) using the shared key and attaches the encrypted validating key in the frame. Before accepting the frame, the recipient decrypts the validating key using the same shared key, looks up the sender's identity token, and validates it with the decrypted validating key. If the validation fails, the recipient drops the frame. The following diagram shows an example of how the AP authenticates an EAPoL-Start request frame using the sender's identity token.

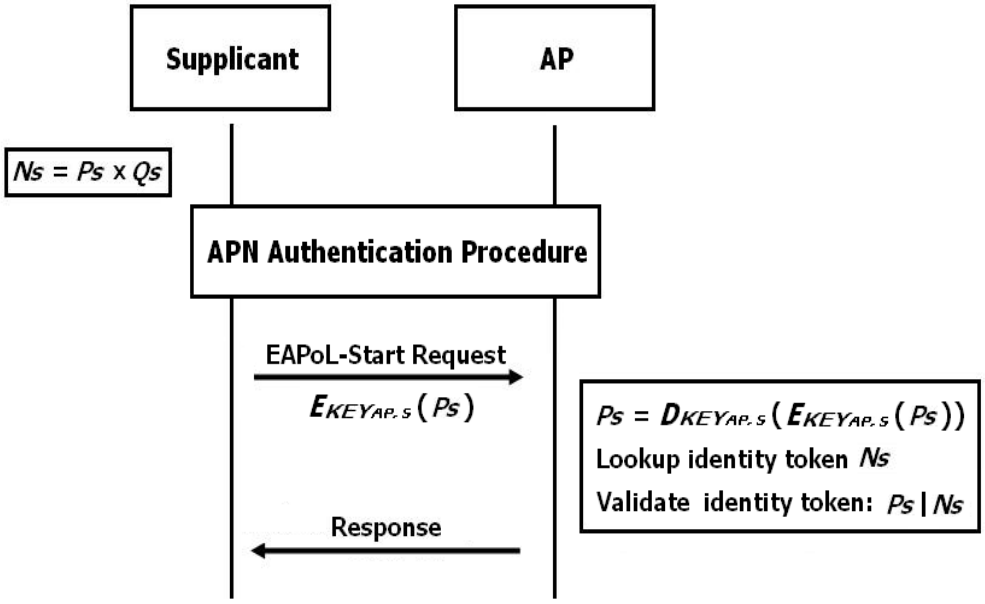

Figure 24: Frame authentication using identity token and validating key

Similarly, an AP can also provide authenticity in its broadcast frames by including its encrypted validating key in the frame. All the recipients upon receipt of a broadcast lookup the previously stored AP identity token and validate it with the decrypted validating key. If successful, the validating key is proven legitimate and the frame is processed by the clients. This will prevent attacks such as a deauthentication DoS attack that attempts to disconnect all the clients in a BSS.

One major difference between authenticating unicast frames and multicast frames is that the validating key in a broadcast frame needs to be encrypted with a group key instead of each individual client's shared key. Fortunately in an 802.11i RSN, all clients and the AP share a single set of group keys produced after a successful group key handshake. Among those group keys, the GTK (Group Temporal Key) allows clients to decrypt multicast and broadcast traffic sent from an AP. By using the GTK[13], an AP can encrypt its validating key in broadcast frames so that all the clients are able to decrypt and validate the request. This provides an effective protection against DoS attacks such as deauthentication or disassociation flooding that spoof a legitimate AP's MAC address.

Although the IEEE 802.11i amendment provides some protection schemes for regular data frames, these protections occur only after a successful EAP/802.1x/RADIUS authentication. Therefore, all the EAPoL frames and partial handshake frames are always out of effective protection and are vulnerable to flooding attacks. In contrast to the 802.11i amendment, the proposed APN authentication scheme accomplishes the EAPoL frame and handshake frame protection by using identity tokens and validating keys.

[13] To be more precise, the Group Encryption Key is used to encrypt the validating key in broadcast frames. Refer to Section 2.4.2.2 for more detail on the Group Key Hierarchy.

5.2.3 Dynamically Refreshed Shared Key

The shared key is dynamically updated for each session to ensure the freshness of the key. Figure 6 in Chapter 2 depicts the RSNA establishment procedure. At the end of the 802.1X/RADIUS authentication phase, both the client and the AS will establish a Master Session Key (MSK). The MSK is bound to the current session between the client and the AS. Using the same computation, both the client and the AS derive a new key, called the Pairwise Master Key (PMK), from the MSK. The AS transfers the PMK to the AP via RADIUS protocol so that both the client and the AP share a common key, which is used to derive a set of other keys (i.e., PTK and GTK) that will be used in protecting a link between the client and the AP. In the proposed APN authentication scheme, the PMK is further used to update the shared key that is used to encrypt an identity token.

Figure 25 describes the update of the shared key using the PMK.

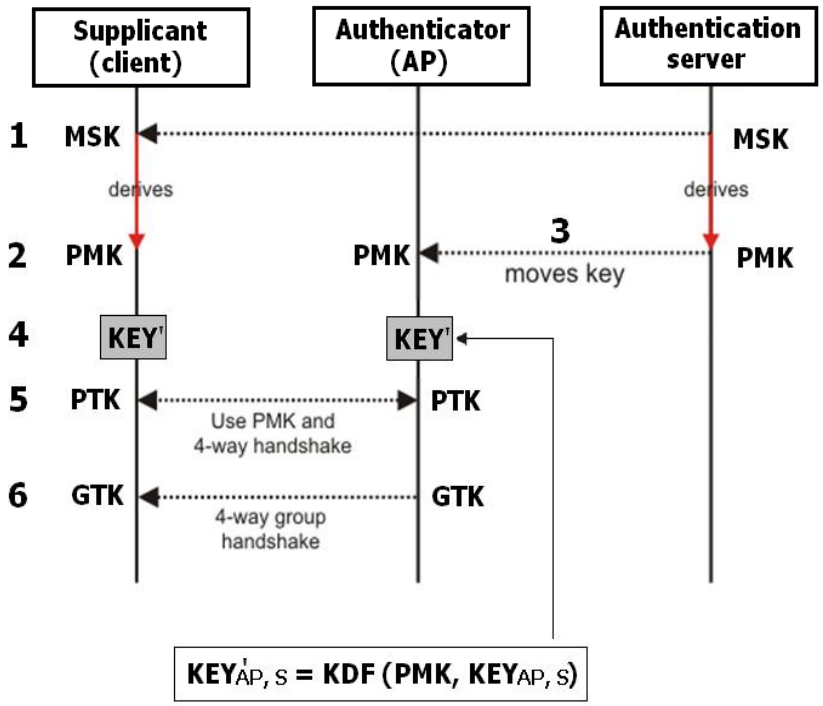

Figure 25: Dynamically updated shared key used in APN authentication scheme

The new shared key ($KEY'_{AP,S}$) is generated by inputting the PMK and the original shared key ($KEY_{AP,S}$) into a key derivation function (KDF). There are several different

types of KDFs that can be used for this purpose [62] depending on the security or performance requirements of the deployment. For the implementation of this research, the PBKDF2 algorithm [63] is used. The PBKDF2 is a SHA1-based KDF for IEEE 802.11i. The main reason for using PBKDF2 is that it has already been used in WPA/WPA2, and the implementation of it is readily available in the Madwifi driver[14].

5.3 Security Analysis of APN Authentication Scheme

The APN authentication scheme provides a security binding between an identity token and a trusted shared key through puzzle solving. Even if an attacker can spoof its MAC address and sniff an identity token of a legitimate client, it still cannot generate a frame that can be accepted by the AP because the validating keys associated with the token cannot be easily determined. Thus, the APN authentication scheme provides a robust method for assuring an identity that cannot be spoofed.

Further, an identity token provides a means of preserving identity over subsequent transactions that cannot be transferred because a token is associated with a shared key. Therefore, identity tokens can be used to validate subsequent requests. Even if an attacker can sniff the identity token of a legitimate client, the associated validating keys cannot be determined (because cracking a prime factorisation is difficult).

Validating an identity token only requires one division operation, which is a trivial computation task. Even under high rate flooding attacks, token verification can be quickly performed and spoofed frames are immediately dropped without consuming a significant amount of AP's resources. Because the identity token of a client and an AP

[14] A minor modification is required in sha1-pbkdf2.c so that the KDF takes the PMK as one of the inputs instead of a SSID. The function is set to hash 4096 times and generate an output of 256 bits.

are exchanged, mutual authentication is achieved. Both the client and the AP can authenticate frames coming from the other party. This prevents attacks such as rogue AP spoofing and main-in-the-middle (MITM) attacks.

The use of shared key also contributes to access control, which is a common practice in an enterprise network environment. Not only the client who has the shared key can access the network, the key is further used for APN authentication to mitigate DoS attacks. Depending on the deployment requirements, the shared key can be a static key that is used as an input to generate a separate shared secret, which is then used for APN authentication and dynamically updated, or used as a dynamic key itself.

On the other hand, the drawback of using a shared key is that extra effort for key distribution and management is required. The security of APN authentication scheme relies on the shared key: if the shared key of a legitimate client is compromised, an attacker will be able to solve the client puzzle in time to establish a session with the AP and launch attacks. For this reason, the shared key is always refreshed every time an 802.1X mutual authentication is completed. This will ensure that the shared key is different for each session and can only be used over a small time span (i.e., from the first authentication response containing the client puzzle until the client receives the EAPOL-SUCCESS message from the AP).

Another advantage of the APN authentication is the fairness of puzzle solving. Unlike traditional client puzzles where legitimate clients are unjustly penalised for proving its authenticity by solving a puzzle, the APN authentication scheme constructs the puzzle in a way that a legitimate client (i.e., one who knows the shared key) requires no effort in solving the puzzle as the solution is the hash digest of the shared key which the client can readily compute.

5.4 Chapter Summary

With the proposed APN authentication scheme, management frames, EAPOL request frames, and any important link-layer information exchanged between a wireless client

and an AP can be authenticated to improve link-layer security. Other control frames or even some data frames can also be authenticated in the same way if necessary.

The design of the APN authentication only requires small number of changes to the original 802.11 implementation, and all the requirements identified earlier in Section 4.3 are achieved with the proposed solution. Due to its stateless and lightweight nature, frame authentication can be performed without causing significant overhead. The APN authentication starts with a stateless authentication exchange to bind a legitimate identity to an identity token, which will be used to authenticate subsequent frames. Both participants are able to prove that all subsequent frames are sent from the party that participated in the APN authentication exchange procedure. This eliminates the most prominent DoS attacks based on injecting or flooding spoofed frames, such as Deauthentication and Disassociation attacks. Therefore, the proposed APN authentication scheme extends the existing RSN to form a much more DoS-resistant and secure wireless network. Further implementation details and experimental results are available in Chapter 8.

Chapter 6

Link-Layer Handoffs in WLANs

The IEEE 802.11i standard provides enhanced security for WLANs by using the RSN architecture. Undoubtedly the increased level of security comes with the penalty of considerable traffic overhead and computation resources which become an obstacle to seamless handoffs. Typically there is a latency of 800 milliseconds to 1300 milliseconds associated with a layer-2 handoff before the mobile STA can resume data transmission through the new AP. Unfortunately, real-time services cannot tolerate such long latency as far as meeting quality of service requirements is concerned. Therefore, supporting continuous mobility and seamless handoff in WLAN has been a challenging issue and a popular research topic.

This research further focuses on link-layer intra-domain (i.e., within the same IP subnet) handoffs and addresses the problem of fast and secure roaming of STAs. The major component that introduces handoff latency is the discovery of new APs and the process of re-authentication to the new AP. Chapter 7 will investigate the AP discovery procedure and devise a solution to reduce the latency of AP discovery during handoff. This chapter is mainly concerned with the re-authentication process and mechanisms that are useful for shortening the latency of the re-authentication process with the new AP.

Several fast handoff mechanisms have been proposed [28, 64-70] by many researchers aiming to keep the handoff re-authentication latency short enough to support real-time multimedia services in WLANs. This research surveyed several existing secure handoff solutions. Their experimental results demonstrated a substantial reduction in the handoff latency. However, very few of these works actually take into account DoS vulnerabilities. Furthermore, they all consume different levels of system resources,

such as producing extra overhead, or requiring heavy computation of the topology of neighbour APs, etc.

In this chapter, the current handoff process will first be examined in detail in Section 6.1. To help optimise the handoff process, various existing techniques and proposed solutions for achieving fast handoff will be studied and discussed in Section 6.2. Based on the analysis presented in this chapter, the goal of this aspect of the research is to investigate a handoff solution that can work in conjunction with the proposed APN authentication as well as providing DoS-resistant seamless handoffs, without sacrificing the level of protection of IEEE 802.11i. The result is a novel fast handoff scheme, which will be proposed and discussed later in Section 6.3.

6.1 Background

One of the major benefits of WLANs is the mobility that allows users to move around while still being connected to the network. However, such access is geographically limited, as users can only move within the radio coverage of an AP while maintaining link-layer connectivity. To retain the connectivity as users move from one AP to another, a handoff process is required that detaches the client from the current AP and attaches it to another, as depicted in Figure 26.

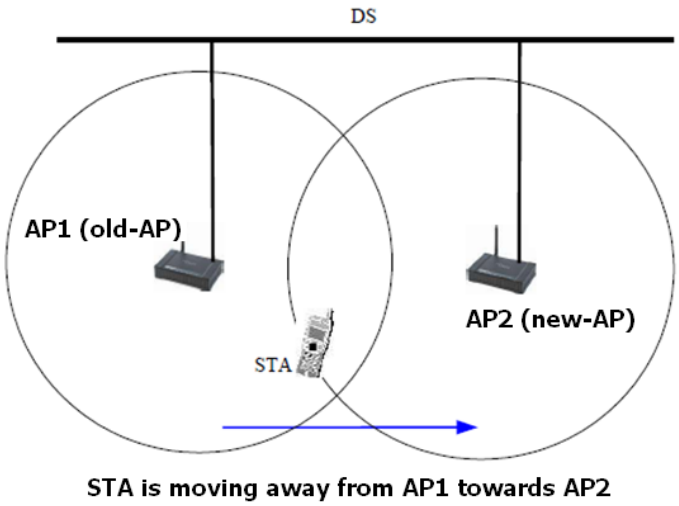

Figure 26: Link-layer handoff within ESS

As a STA moves and loses connectivity with its current AP, it starts a handoff process. This requires several necessary operations, including: (1) discovery - identifying a suitable candidate AP for handoff, (2) IEEE 802.11 authentication - Open System Authentication is used here for 802.11i, (3) (re)association with the candidate AP, and (4) IEEE 802.11i authentication – authentication using 802.1X/EAP upper layer method and key derivation with four-way handshake. The message flow of a complete handoff process in an 802.11i secured WLAN is summarised in Figure 27.

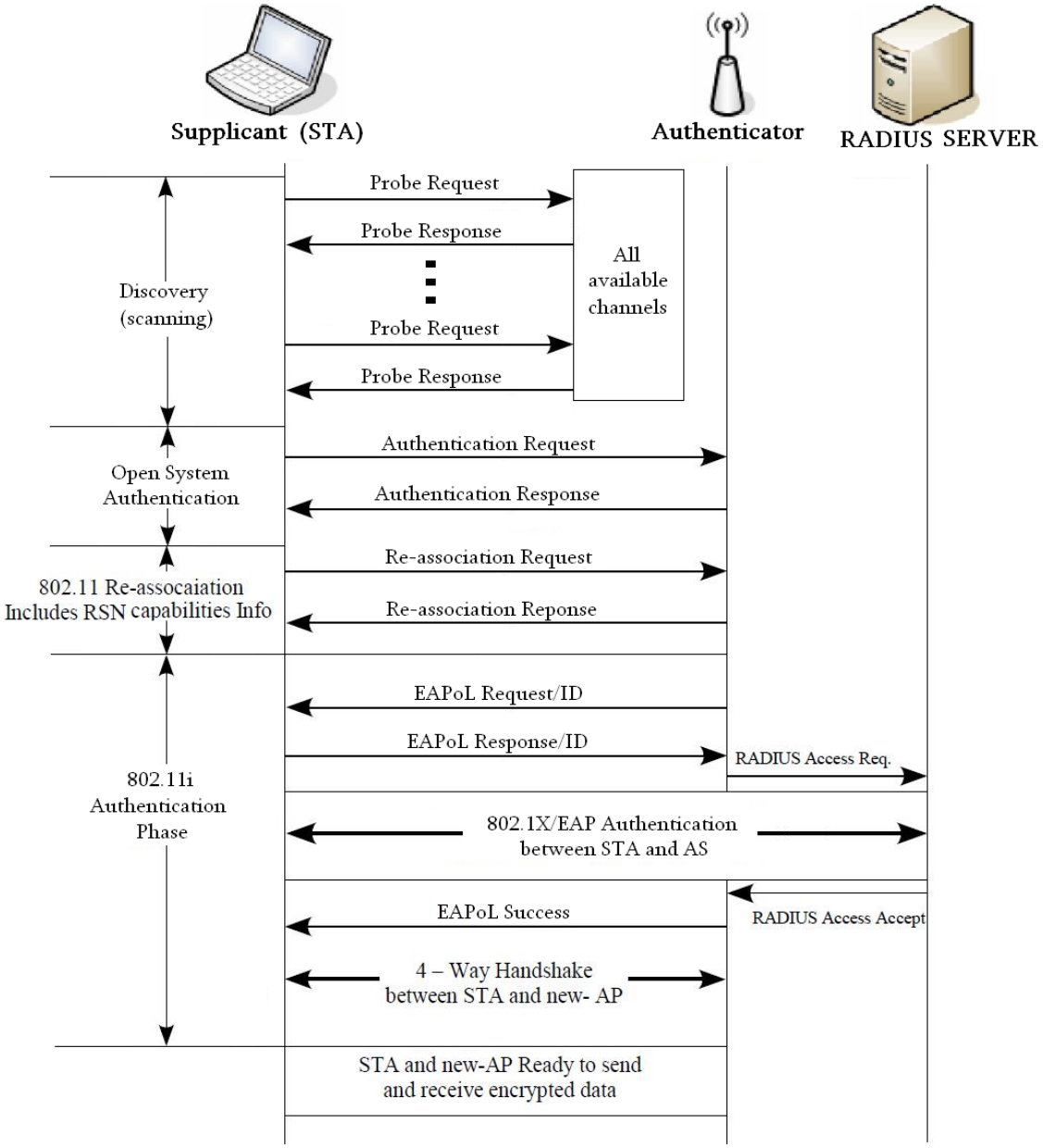

Figure 27: Message flow during handoff in 802.11i secured WLAN

This current handoff process involves a large latency, which can range from several hundred milliseconds (ms) to almost 1.5 seconds, before the STA can resume the secured data communication through the new AP. The discovery phase and the 802.11i re-authentication phase are the main sources of handoff latency, because they require a long time to complete. In the discovery phase, the STA learns which available APs are present in vicinity by broadcasting probe requests (active scanning) on all the available channels. The full scanning usually takes about 300ms to 500ms to complete. From the received probe responses, the client selects a new candidate AP based on some implementation-dependent policy[15] and proceeds with the handoff. The 802.11i re-authentication is used to re-establish key materials, such as PTK and GTK, between the STA and the new AP, so that encrypted data transmission can then run. The 802.11i re-authentication latencies have been reported to be of the order of 1 second [71]. This long handoff latency becomes the major obstacle to delivering reliable VoIP service over WLAN.

Furthermore, this current model of 802.11i re-association mechanism can be exposed to many kinds of attacks such as unauthorised disassociation or fake handoffs [8, 72] as well as flooding attacks described in Chapter 3. This is because the current standard does not describe specific security policies and algorithms related to the re-association control in AP. The management frames involved in the handoff are also not protected.

Supporting voice and multimedia with continuous mobility requires the handoff latency to be small. Specifically, the overall latency should not exceed 50 ms to prevent excessive jitter and quality degradation [73]. Therefore, meeting the high expectation of voice service over WLAN requires the support of secure and seamless handoff of a STA. To achieve this without compromising the level of security is another major focus of this research.

[15] Chapter 7 will propose a network-based mechanism that will help a client select the best candidate AP without performing a full active scanning so that handoff delay is further reduced.

6.2 Classification of Related Works

This section provides an overview of some of the existing link-layer handoff schemes for handoff in WLANs. Most of them are designed to shorten the 802.11i re-authentication delay by the means of pre-distributed keys or predicting STA movement. Although fast re-authentication is generally attainable in practice, the remaining challenge is how to achieve fast re-authentication at a low cost without compromising the 802.11i level of security.

In this research, the handoff schemes are roughly classified into five categories: (1) full authentication, (2) full authentication with lightweight EAP method, (3) proactive key distribution (PKD), (4) pre-authentication, and (5) IEEE 802.11r Roaming.

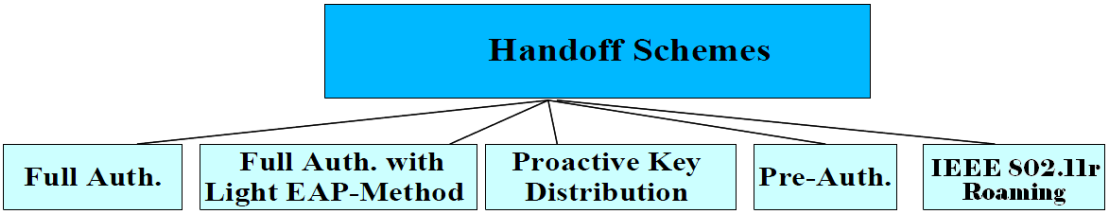

Figure 28: Classification of existing handoff schemes

The first category is the simplest way of performing re-authentication which is to repeat the same full 802.1X/EAP authentication with the new AP. This approach incurs long latency and high communication overhead. The second category [66, 74] also involves the full 802.1X authentication, but instead of using secure EAP methods (such as EAP-TLS), the re-authentication uses lightweight EAP methods such as EAP-MD5. The re-authentication latency is reduced because less communication overhead is needed. However, for environments where the APs are physically located at a distance from the AS, the authentication process still incurs long latency. This is due to the fact that the communication latency between the AP and the AS contributes most to the overall authentication-related latency.

The remaining three categories of handoff scheme will be discussed in more detail in the following sections.

6.2.1 Proactive Key Distribution

The Proactive Key Distribution (PKD) is another popular technique that aims to achieve fast handoff by shortening the re-authentication delay. Several PKD based solutions [16, 65, 68, 70, 75-77] have been proposed. The principle behind these solutions is to allow the AS (or the old AP) to proactively distribute the PMKs to potential target APs so that the re-authentication will not require the AS's involvement when the STA roams to the new AP.

The entity (AP or AS) which distributes the PMK is required to maintain a database of users' roaming patterns in order to facilitate some prediction schemes [68, 70, 75, 76] that select potential next APs to pre-distribute keys to. The most popular prediction scheme used for PKD based handoffs is the concept of the Neighbour Graph, which is introduced in [75]. The Neighbour Graph is a data structure which represents the current network topology. The Neighbour Graph can be constructed in a distributed manner at each AP or it can be installed on the AS when WLAN is deployed. The latter is more often used because of the faster convergence time.

By using a Neighbour Graph, the STAs' contextual information (such as the session key derived by both the STA and the AP) are proactively distributed to all the adjacent APs in advance. If a STA moves to one of the candidate APs in the Neighbour Graph, the authentication process can be avoided.

These PKD schemes have the following drawbacks. First of all, the performance of the scheme depends on the cell environment. With high STA density AP coverage, it is likely that the context of a particular STA (STA*i*) in a candidate AP could be updated by the other STAs before STA*i* handoffs to the AP. If the security context is not found when STA*i* hands off, a full authentication process is still required. Secondly, PKD schemes require state information to be maintained in a centralised AS. If there is no proper protection, the AS can become vulnerable to DoS attacks. Lastly, key distributions require extra signalling and computation overhead between the distributor and the neighbouring APs and the cost of periodically running the prediction algorithms. If the number of STAs grows and/or STAs roam frequently, the

computational overhead will become a performance bottleneck. Therefore, PKD based fast handoff solutions are usually non-scalable. Solutions such as [75] feature lower computational overhead by caching context between APs, but they suffer from weak security due to the reuse of PMKs among APs.

6.2.1.1 IEEE 802.11F: IAPP

Some standard efforts are being carried out that aim to provide solutions for fast and secure roaming. IEEE 802.11f or Inter-Access Point Protocol (IAPP) [78] is a recommendation that describes an optional extension to IEEE 802.11 that will enable multi-vendor AP interoperability within the DS. IAPP is used by some of the PKD schemes [65, 68, 77] to distribute security keys because it provides a mechanism that allows the transfer of the STA related context (such as PMK) from the previous AP to the new one upon handoffs. It also defines a cache mechanism that allows APs to exchange this information pro-actively (i.e., before handoff) or after the STA hands off to the adjacent new AP.

- ***CONTEXT TRANSFER***

IAPP provides a standard way for one AP to move a STA's context to another AP. This feature is useful because security contexts such as the 802.11i PMK, PTK, GTK, etc may be transferred to the new AP to decrease handoff latency. With IAPP, the AP transfers the STA context to its AP neighbours using a Cache-Notify message. Each neighbour responds with a Cache-Response message in order to confirm its cache has been updated. To secure IAPP exchanges between APs, IEEE 802.11f defines the use of the RADIUS protocol to ensure the confidentiality of the context transferred over the distribution system [78].

- ***SINGLE ASSOCIATION***

Another design goal of IAPP is to enforce the single association requirement. The IEEE 802.11 requires that the STA should only associate with one AP at any moment. For handoffs using the existing 802.11 standard, the STA indicates an association

change by notifying the old AP with a disassociation frame. If this disassociation frame is lost, the old AP may still think that the STA is associated with it. This becomes a situation where the STA is actually associated to multiple APs. Multiple association relationships will confuse the layer-2 switches that connect the APs because the switches have no way of knowing where to forward the frames to reach the STA.

IAPP prevents this problem by enforcing the single point of association by using some notification messages. When a STA is associated to an AP, the AP broadcasts an Add-Notify packet to all other APs, notifying the STA's current association. Upon receiving an Add-Notify packet, the other APs remove the association state for the specific STA.

To handoff to the new AP, the STA will use a re-association frame to notify the new AP about the MAC address of its old AP. This indicates the STA is changing its association to the new AP. The new AP consults the AS to get the IP address of the old AP based on the MAC address of the old AP. The new AP can then use IAPP to communicate with the old AP. Upon receiving the IAPP notification from the new AP, the old AP will remove the association state of the STA. The new AP also broadcasts layer-2 update frames on behalf of the newly associated STA for layer-2 routing update.

■ ***IAPP MESSAGES***

This standard defines a set of messages for communication between APs:

- ADD-notify: a multicast advertising packet addressed to the APs in the same subnet to notify the association of a node with the advertising AP. This packet must be protected.
- MOVE-notify: this is an unicast message sent, in a reactive way, from the new AP to the old AP to request security information (context).
- MOVE-response: this is the answer message by the old AP. It includes context information related to the re-associating station like a PMK.

- Send-Security-Block and ACK-Security-Block: is a two-message exchange between the old AP and the new AP to setup a security association for securing the inter-access points exchange.

- CACHE-notify: this message invites the neighbouring APs to proactively cache the security context of the mobile node. This message is used and extended by the handoff scheme proposed in Section 6.3.

- CACHE-response: this is an optional packet, sent as a response to the CACHE-notify to advertise that the context information is present in the cache. The proposed handoff scheme will further extend this message to include additional security material.

In summary, IAPP allows a new AP to obtain the existing PMK of a STA from the old AP, or it can be used to distribute existing PMKs to adjacent APs before the handoff occurs. This scheme can involve certain vulnerabilities associated with the fact that the PMKs are unnecessarily shared among neighbouring APs. Hence, any particular AP has access to the PMK information of all STAs associated with adjacent APs, even though those STAs may never associate with the AP.

6.2.2 IEEE 802.11i Pre-authentication

The full 802.1X authentication can take up to several hundred milliseconds to a second to complete, depending on the chosen EAP method. In a roaming scenario, the STA has to be authenticated by every AP it wishes to attach to, in order to gain network access. This results in a significant handoff latency and packet loss. In scenarios where STAs often move back and forth between APs, the repetition of full 802.11i authentication can also cause severe network performance degradation. To reduce the link layer handoff latency, the IEEE 802.11i [4] proposed two options to minimize the latency: pre-authentication and PMK caching. The following sections will briefly describe those mechanisms.

6.2.2.1 Handoff with Pre-authentication

Pre-authentication can also be referred to as “association in advance”, which allows the STA to communicate with and pre-authenticate to APs to which the STA is expected to handoff to via the wired backbone DS connection through the currently associated AP. Unlike PKD based schemes, pre-authentication does not require any mobility prediction mechanisms. The pre-authentication is initiated by the STA and the selection of candidate APs for the pre-authentication is the STA’s decision based on APs’ capabilities and its availability in the radio range.

To initiate the pre-authentication, the STA sends, via the currently associated AP, an EAPOL-Start message with the BSSID of the new AP attached. At the end of the pre-authentication process, both the station and the new AP will generate the same pairwise master key (PMK) which will be cached. If the authentication is successful, the newly generated PMK will be cached on the target AP, thus establishing a PMK Security Association (PMKSA) between the STA and the target AP. Later when the STA roams to the new AP, the STA sends to the AP a reassociation request with a PMK identifier (PMKID) attached. The PMKID is a 16 byte token to identify the cached PMK. If the new AP does in fact have the PMK associated with this PMKID, the cached PMK will be used to derive new session keys directly between the STA and the new AP with a four-way handshake. Using the pre-cached PMK instead of 802.1X full authentication for the re-authentication of the handoff STA reduces handoff latency significantly. The 802.11i pre-authentication procedure is depicted in Figure 29.

The first time a STA associates with an AP in the network, a full 802.11i authentication is still required. However, if the STA knows where it will be roaming, the STA can pre-authenticate (doing full 802.11i authentication but without the four-way handshake) to a new target AP using the current AP via the backbone DS.

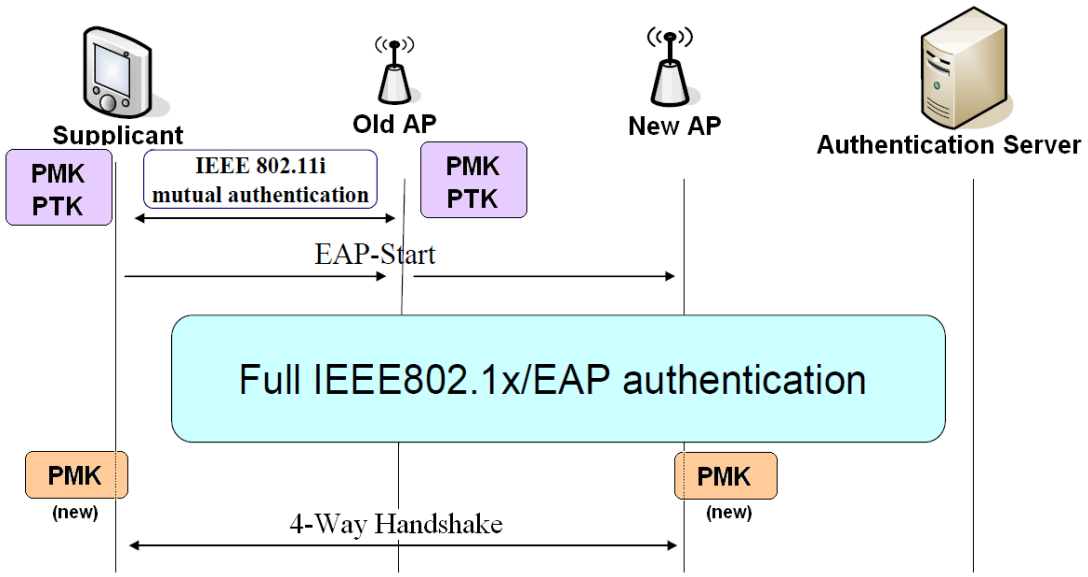

Figure 29: IEEE 802.11i Pre-authentication

IEEE 802.11i pre-authentication provides a way to establish a PMKSA with a new AP before the STA associates. Faster handoff could be achieved as a result. However, pre-authentication also has some limitations:

1. Pre-authentication requires more processing resources at APs and at authentication servers because STAs performing pre-authentication will create significant overhead and extra load on the authentication server.
2. Full EAP authentication is still required for pre-authentication. This poses a lot of signalling with the authentication server during each movement. Also, the full four-way handshake is needed to complete the handoff when the STA moves to the new AP.
3. The pre-authentication process is as vulnerable to DoS attacks as is the normal 802.11i authentication. The request messages are not protected and the protocol involves unprotected stateful operations.

6.2.2.2 PMK Caching

The basic concept of PMK caching (also referred to as “fast roam back”) is to allow AP and STA to store previous security associations (i.e., the PMK) in their memory as

STA roams away from an AP. When the STA roams back to this AP again, the PMK security association (PMKSA) that is previously established can be used straight away. Because the PMK is already cached, only a four-way handshake is required for the STA to re-associate to the AP. The four-way handshake confirms that the STA and AP have the same PMK security association.

The PMK caching prevents a STA from repeating the entire authentication process and allows for fast re-association with the corresponding AP by merely renewing the PTK out of the PMK via another four-way handshake. Hence, not only is the authentication server's load reduced during roaming back, the handoff latency is also minimised as a result.

6.2.3 IEEE 802.11r Roaming Scheme

Though 802.11f solves the problem for APs to exchange information about the roaming of STAs, it does not solve the problem of speeding up the process of discovering APs, authentication, and association. To further improve the roaming capability in WLANs, the IEEE Task Group proposed a new fast roaming standard, 802.11r [79], attempting to minimize the handoff latency for BSS transition process to support real-time applications.

Like the 802.11i pre-authentication, the 802.11r carries out most of the authentication process before the STA actually starts roaming. The full 802.1X/EAP authentication is first performed to generate the PMK when the STA initially joins the network. After that, a set of keys are derived from the PMK which correspond to each authenticated STA and are distributed to all authenticated APs in the subnet. When a STA roams to a new AP, it finds its corresponding authentication key so that the overhead of a complete authentication process is prevented. In addition, IEEE 802.11r allows a STA to perform part of the four-way handshake and some resource reservation at the target AP, before STA roams. When the STA roams to the new AP, it only needs to re-associate with the target AP to complete the movement.

The IEEE 802.11r standard uses a three-tier key hierarchy. The initial key is the Master Session Key (MSK), which is produced from the 802.1X/EAP authentication. In addition, two levels of Key Holders (KHs) are introduced: (1) level 0 KH (R0-KH) for storing the top level keys PMK-R0s, and (2) level 1 KH (R1-KH) for storing the second-level keys PMK-R1s. PTK is the third level of the key hierarchy and is stored at APs. The R0-KH is the AP that the STA initially associate with during its initial association and shares the secret PMK-R0 key with the STA. KHs are logical entities that can be separate physical entities or can be located within APs. Figure 32 below illustrates the 802.11r key hierarchy.

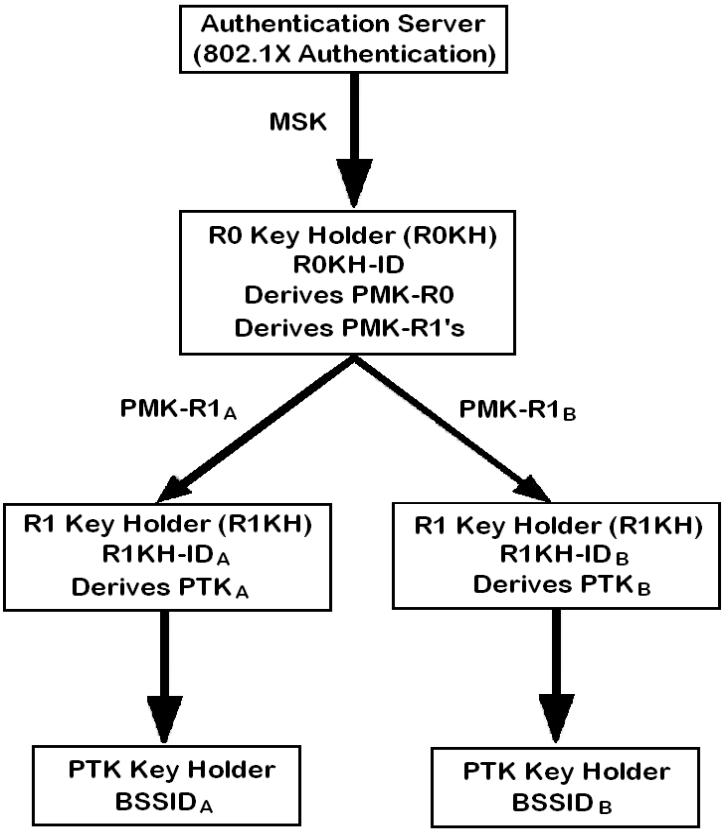

Figure 30: IEEE 802.11r Key Hierarchy

As the figure shows, after the initial 802.1X authentication, the PMK-R0 is mutually derived by the STA and the R0KH from the last 32 octets of the MSK. After that, PMK-R1 is mutually derived by R0KH and the STA and is distributed to all R1KHs within the same subnet from R0KH. Finally, PTK is derived by R1KH and the STA based on PMK-R1.

When a STA requires a handoff, it initiates the FT protocol by signalling a request to the current associated AP which is either an R0KH or an R1KH. This then forwards the request to the R0KH if it is an R1KH. Upon receiving the request, the R0KH derives PMK-R1 from the PMK-R0 and distributes the PMK-R1 to the target AP (R1KH). When the STA roams to the new AP, the PMK-R1 associated with this STA should already exist, so there is no 802.1X/EAP authentication required. The STA and the new AP can then perform the four-way handshake to derive PTKs.

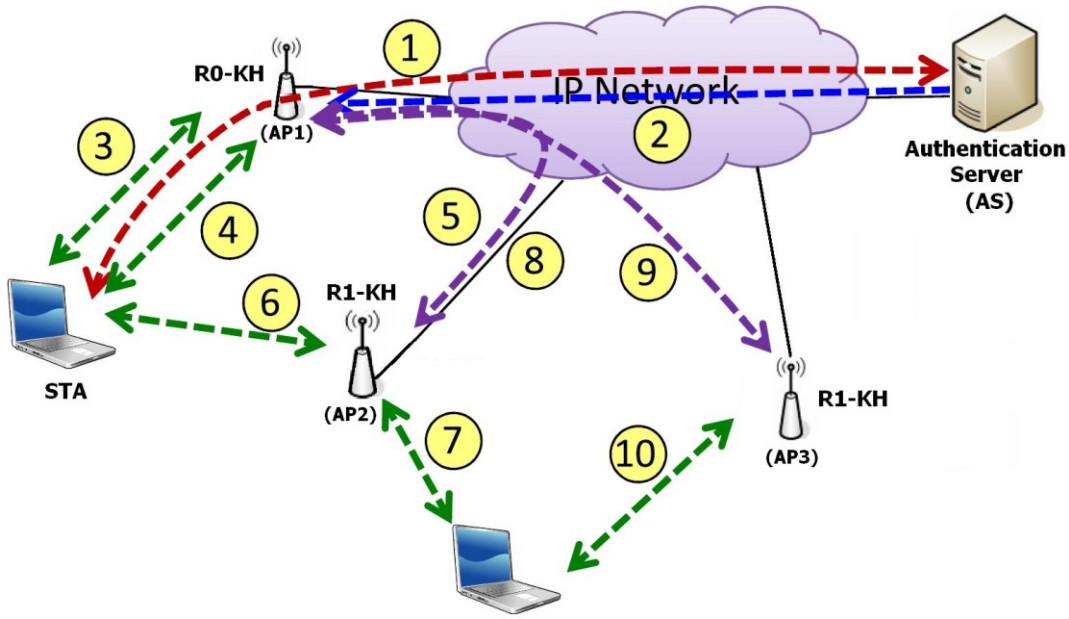

Figure 31: Message exchange for IEEE 802.11r-based handoffs between APs

Figure 31 illustrates the message exchange for handoffs between APs using 802.11r:

1. STA and AS performs 802.1X/EAP authentication when first joins the network.
2. MK transferred from AS to AP1 (R0-KH).
3. Four-way handshake between STA and AP1 and derive PTK.
4. STA signals AP1 to request handoff to AP2 (R1-KH).
5. AP1 generates session key for AP2 and transfers it.
6. STA completes handoff by executing four-way handshake and derives new PTK.
7. STA signals AP2 to request handoff to AP3 (R1-KH).
8. AP2 asks AP1 to generate session key for AP3.
9. AP1 generates session key for AP3 and transfers it.
10. STA completes handoff by executing four-way handshake and derives new PTK.

A performance study [80] was carried out by Intel to evaluate the performance of an 802.11r prototype compared with the 802.11i standard. It was shown that when using 802.11r in roaming scenarios, handoffs involve shorter transition time and reduced packet loss, thereby allowing improved VoIP voice quality. Table 3 from [80] gives an indication of the average roaming time and the packet loss using 802.11i pre-authentication and 802.11r fast transition mechanism.

Authentication approach	Average roaming time	Average packet loss %
Basic 802.11i	525 ms	1.8
802.11r fast transition	42 ms	0.2

Table 3: Handoff performance comparison between 802.11i and 802.11r

Although 802.11r seems promising in roaming scenarios compared to the 802.11i pre-authentication scheme, there are still a few shortcomings with this scheme. Firstly, distributing STAs' authentication keys among APs belonging to the same subnet does not consider the overall network performance. A mechanism that could help the client determine the best target AP(s) is needed to complement 802.11r, otherwise distributing keys to all available APs in the ESS is not a scalable solution in dense deployment scenarios.

Secondly, in 802.11r architecture the security relies on the reliability of R0-KH. The R0-KH may actually be located close to the edge, thereby creating the vulnerability that if R0-KH is compromised, all PMK-R1s derived from the corresponding PMK-R0s will also be compromised.

Moreover, Clancy [64] claims that R0KH, which is responsible for deriving and delivering the PMK-R1 keys to new APs, could be an AP on the wall, and is considered untrusted in the 802.11i security model. This could be a security weakness that downgrades the overall security level of 802.11r as described in [81].

6.3 Proposed Fast Handoff Scheme

In this section, a novel seamless and secure handoff scheme called Fast AP Transition Protocol (FATP) is proposed. The existing IEEE 802.11i specification is extended to support the proposed handoff scheme for backward compatibility.

The proposed FATP scheme uses the proactive key distribution technique to transfer existing security context to a handoff candidate AP prior to handoff. It further allows a roaming STA and the candidate AP to mutually verify the identity of each other and derive new session keys without the involvement of the AS upon STA re-association with the new AP.

With traditional pre-authentication schemes, a full 802.1X/EAP authentication is required to re-establish the trust relationship between the roaming STA and the new AP during handoff. Although the re-authentication latency is reduced as a result of doing pre-authentication prior to the disconnection of the STA, the STA will still suffer from poor handoff performance because the ongoing data communications is disrupted while the STA is performing the full 802.1X authentication with the new AP. In other words, the time spent in the mutual authentication with the new AP prior to handoff would still contribute to the service outage time that should be minimised. On the other hand, with the proposed fast handoff scheme, there is no need to go through a full 802.1X authentication to perform a handoff and the involvement of the AS is not required to re-establish a trusted relationship. Therefore, the overall handoff delay and packet loss can be significantly reduced to achieve seamless and secure roaming.

The following sections 6.3.1 and 6.3.2 will discuss the requirements, the trust model involved, and trust associations between APs in the FATP scheme. The FATP handoff scheme will be described in detail in Section 6.3.3. Security analysis of the scheme and more discussion will be given towards the end of the chapter in Section 6.4.

6.3.1 Requirements

The FATP handoff scheme provides a means to securely transfer the existing trust relationship between the roaming STA and the current AP to a new AP prior to handoff. When the STA roams to the new AP, only a verification of the trust relationship is needed before secured data communications can resume through the new AP. Transferring the trust relationship instead of re-establishing it prevents the considerable 802.1X authentication overhead and the long transmission latency associated with the backend AS. The transferred trust relationship will allow the new AP and the roaming STA to mutually verify each other's identity based on the previous security context obtained from the initial full 802.1X authentication. Because the AS is not involved in the re-authentication, the handoff latency is reduced.

To achieve a secure trust relationship transfer between APs, the following two issues need to be considered:

1. Trust authority delegation from the AS to APs:

For the new AP to be able to verify the identity of a roaming STA without involving the 802.1X authentication with the backend AS, authority must be delegated to the AP so that it can authenticate the STA on behalf of the AS. To do this, the 802.11r key hierarchy concept is adopted in the proposed FATP handoff scheme.

When the STA first connects to the network, a full 802.1X authentication is performed so that the STA and the associated AP (with the role of R0-KH) share an initial PMK (PMK-R0). Because this AP maintains the MK received from the trusted AS, it is considered to have the trust authority to verify the STA. As the STA decides to handoff to a new AP (with the role of R1-KH), the current AP delegates the existing authority to the new AP so that the new AP can authenticate the roaming STA without having to obtain authority from the AS through 802.1X authentication. In the 802.11r context, the current AP delegates the authority by generating a new PMK (PMK-R1) and delivering it to the new AP.

2. Secure L2 communication channel between legitimate APs:

To transfer PMK-R1s from the R0-KH (i.e., the initial AP) to other APs, a trust relationship must exist between the APs so that keying material can be delivered in a secure fashion. However, the current 802.11r standard does not specify how to transfer this information from the R0-KH to R1-KHs.

Since the IEEE 802.11f (IAPP) provides a standardised way of exchanging information between APs located within the same subnet, the FATP makes use of its capability for pre-distributing security context before handoff. To secure the IAPP exchanges between APs, 802.11f defines the use of the RADIUS protocol to provide APs' mutual authentications and ensure confidentiality of the security context transferred over the distribution system [78].

6.3.2 Handoff Trust Model

The IEEE 802.11i standard is based on the following assumptions with respect to trust relationships: (a) a STA trusts the AS and the AP with which the STA is associated, and, (b) the STA does not trust all non-associated APs in the first instance. Figure 32 depicts the trust relationships based on the 802.11i standard. The solid arrows represent an explicit mutual trust relationship while the dotted line represents an implicit trust relationship that must be created in order to make security claims about the communications path [4].

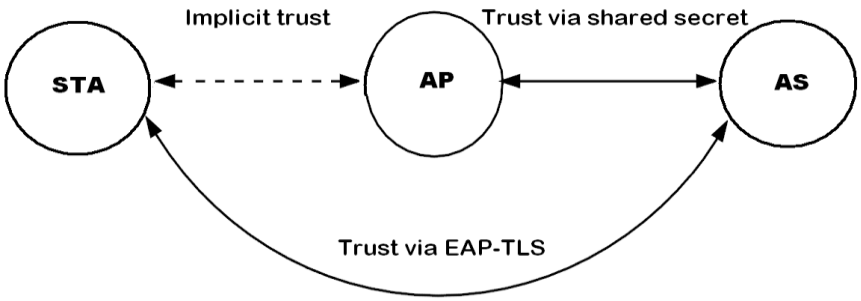

Figure 32: The 802.11i Trust Relationship

Under the 802.11i standard, all the legitimate APs share a common secret key with the AS. Therefore, there is a direct (or explicit) trust relationship between an AP and the AS. To establish a trust relationship between a STA and an AP, they must mutually authenticate each other and derive common keys based on keying material from the mutually trusted AS. This will require an 802.1X/EAP authentication. Therefore, the trust relationship between a STA and an AP is implicit, as it is derived from the EAP-TLS authentication that binds the STA and the AP into trust via the trusts AS.

Based on this existing trust relationship, Figure 33 depicts the trust model that the proposed handoff scheme is based upon.

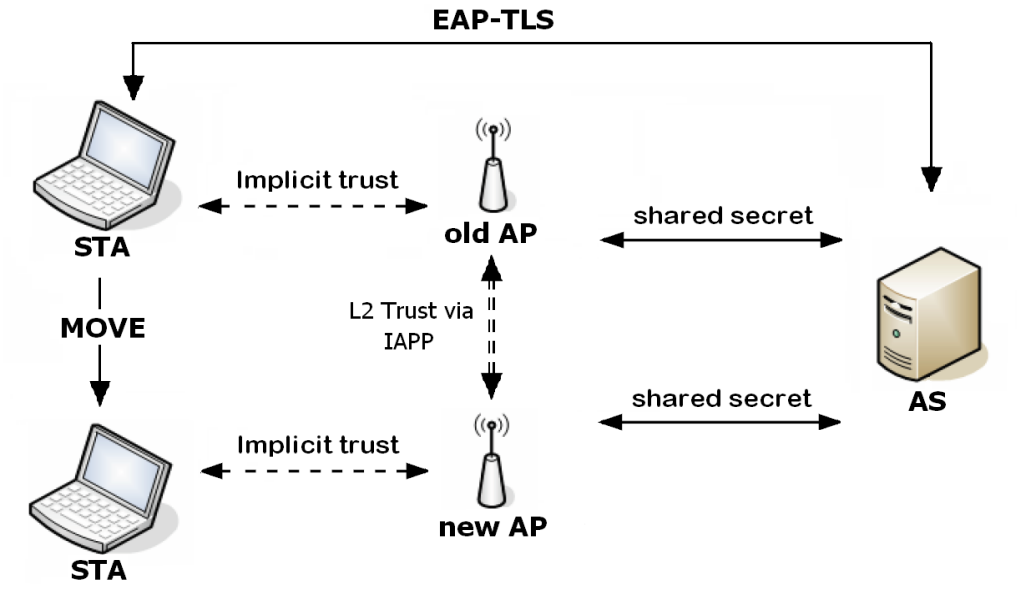

Figure 33: FATP Handoff Scheme Trust Model

Before handoff, the trust relationship between the STA and the (old) AP is established based on the 802.1X (EAP-TLS) authentication. As the STA moves towards the new AP, a trust relationship needs to be established with the new AP. The most straightforward way of doing this is to repeat the same 802.1X authentication with the AS so that the new implicit trust is created from scratch. This approach, however, incurs long delays and is inefficient because the existing explicit and implicit trust relationships are not fully exploited.

To achieve a fast handoff, the FATP scheme provides a means to proactively transfer the existing trust relationship of the old AP to the new AP in order to prevent the need of an 802.1X authentication. The IAPP protocol is utilised to provide the required secure link-layer communication between APs.

In the existing 802.11i standard, a handoff is considered the same as the initial connection to the network, which will require an establishment of session dependent key materials to derive new session keys. However, in the research it is argued that handoff is effectively the continuation of the on-going session because the STA could continue using the same session key if the STA does not handoff. Because in an enterprise WLAN environment where trust relationship between APs can be established by means such as the IAPP protocol, the recreation of keys from scratch for a handoff session becomes unnecessary, provided that the freshness of session keys and the liveness of the communicating entities will not be sacrificed[16]. Therefore, it can still be considered secure for STAs and APs to use existing key materials to derive new session keys for re-authentication, as long as the session keys are different for each session (either a new connection or a handoff session). Based on this consideration, the re-authentication process can be simplified to a key refresh process, which does not require a full 802.1X process. As a result, the re-authentication latency can be significantly reduced without compromising the original 802.11i security level, provided that security keys can be reliably transferred between APs. In this research, it is assumed that the use of IAPP for inter-AP communications can deliver this guarantee[17].

[16] This claim is only valid in an enterprise WLAN environment. In a public WLAN, such as hotspots, APs normally do not have a secret key that is shared with the AS. Therefore, it is not possible to establish a secured channel between APs unless there exists some shared secret.

[17] IAPP defines the use of RADIUS protocol to distribute keys to APs for encrypting information transferred in the DS, therefore confidentiality of the information exchanged is ensured.

6.3.3 Fast AP Transition Protocol (FATP)

This section presents the proposed FATP scheme which provides secure and seamless intra-domain link-layer handoffs. The FATP handoff involves a mixture of PKD and pre-authentication techniques. The STA is able to request a pre-distribution of security keys and context transfer to the target AP via the current associated AP prior to handoff. The resulting trust relationship with the new AP will still have the same properties of a full EAP/TLS authentication, but at a significantly less cost in terms of latency and computational power of the STA and traffic overhead in the network.

The FATP implements the 802.11i protocol paradigm: authentication (i.e., to establish a trust relationship) first, and then re-association (i.e., to change AP attachment). Based on this paradigm, the FATP scheme comprises two stages: trust transfer prior to handoff and fast re-association after moving to the new AP. Figure 34 depicts the top level concept of the FATP scheme.

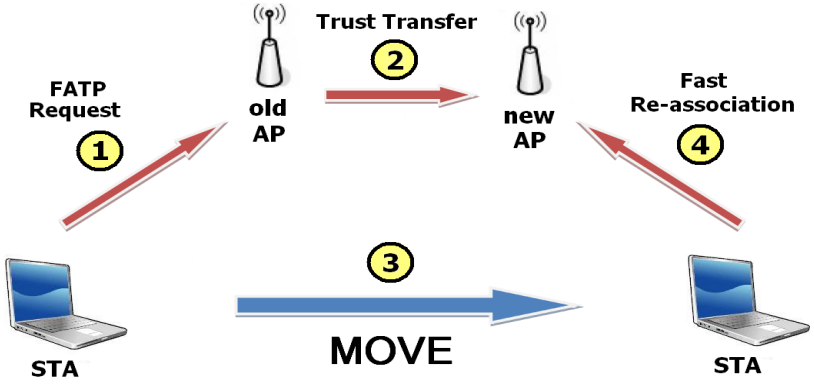

Figure 34: Top level concept of the proposed FATP handoff scheme

The trust transfer stage takes place before handoff to transfer the existing security context from the initial AP (which authenticated the STA with 802.1X authentication) to the new AP. After the STA moves to the new AP, the fast re-association stage completes the handoff by updating the point of attachment of the STA to the network. With this model, outage periods during handoffs are reduced to the time it takes to perform re-associations and network configurations, but not authentications. In addition, based on the previously exchanged identity tokens, the messages involved will be authenticated, preventing DoS attacks that are not addressed by most of the existing handoff solutions.

6.3.3.1 Notations

The following table defines the notations used in the thesis to describe the proposed FATP handoff scheme.

Notation	Meaning
R0-AP	The AP that performed an 802.1X/EAP authentication with the STA. It is responsible for delegating trust and authority by generating and transferring PMK-R1s to R1-APs upon request.
R1-AP	An AP to which the STA wants handoff. It does not have the trust and authority to authenticate the STA until it obtains a PMK-R1from the R0-AP of the STA concerned.
PMK-R0	The Pairwise Master Key (PMK) generated as a result of a successful 802.1X/EAP authentication. This is shared by the STA and the R0-AP that authenticated the STA
PMK-R1	A new PMK derived from an existing PMK-R0 by a R0-AP, and is to be delivered to the requesting R1-AP. The R1-AP that possesses a valid PMK-R1 is considered trusted and authorised to be associated by the roaming STA.
N_S'	A STA's new identity token (as defined in Section 5.1.3).
N_A	An AP's identity token (as defined in Section 5.1.3).
SNonce	A random bitstring generated by a STA.
ANonce	A random bitstring generated by an AP.

Table 4: Notations used for the FATP handoff scheme

6.3.3.2 Context Transfer

The FATP scheme uses the IAPP standardised messages to transfer the security context and authentication material (such as the STAs' PMK-R1s) between APs in a proactive way to reduce handoff latency. The FATP defines two types of intra-domain handoff scenarios, namely "R0 to R1" handoff and "R1 to R1" handoff.

Figure 35 depicts the two possible FATP handoff scenarios. In the R0 to R1 handoff, the STA is roaming from a R0-AP to a R1-AP (e.g., the R1-AP1 in the figure). Similarly in the R1 to R1 handoff, the STA is roaming from a R1-AP to another R1-AP in the same network (e.g., from R1-AP1 to R1-AP2 in the figure).

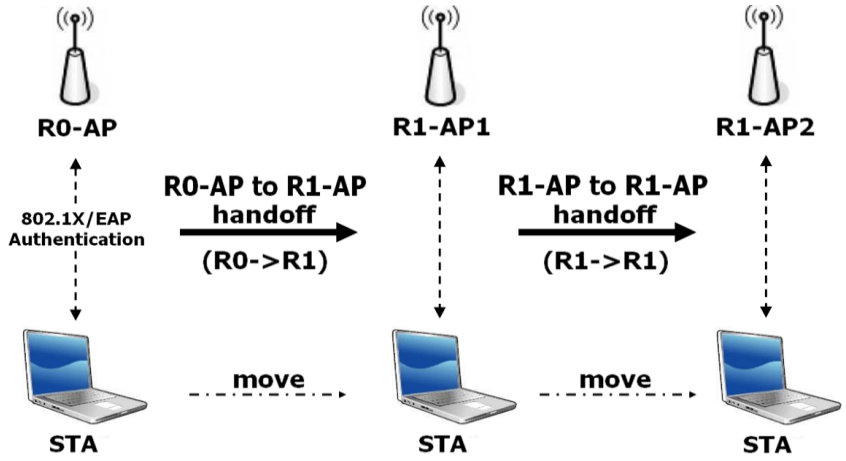

Figure 35: The two FATP handoff scenarios

For the R0→R1 handoff, the current AP is a R0-AP, which already knows the PMK-R0, and therefore the R0-AP is able to directly generate a PMK-R1 and delivers it to the target AP. Two existing IAPP messages, namely Cache-Notify and Cache-Response, are used to transfer the PMK-R1 and related security context from the R0-AP to the target R1-AP.

On the other hand, for the R1→R1 handoff, the current AP does not know the PMK-R0 of the STA, therefore the current AP needs to request the STA's R0-AP to generate one and transfer over. Two new IAPP messages, PMKR1-Request and PMKR1-Response, are introduced for this purpose. Once the current AP obtains a PMK-R1 from the STA's R0-AP, it transfers the PMK-R1 to the target AP using the same IAPP cache method used in the R0→R1 handoff scenario.

Those IAPP messages are IP packets carried in a TCP session between APs, therefore a mapping between the BSSID and the IP address of the target AP is required. The 802.11f defines the use of RADIUS to provide such a mapping. In addition, the RADIUS also provides security blocks for both the current and target APs. The security blocks each contain information generated dynamically by the RADIUS server

to be used for securing the AP to AP communications. After exchanging security blocks, both APs will have the information needed to encrypt all subsequent packets transferred between the APs in the session[18].

The target AP can then use the received PMK-R1 (either directly from a R0-AP or through another R1-AP, depending on the handoff scenarios) and generate new session keys without going through the full 802.1X authentication process or requiring the involvement of the AS.

6.3.3.3 FATP Trust Transfer Procedure: R0-to-R1 handoff

Prior to handoff, the STA uses FATP to request a trust transfer to the target AP specified by the STA. Figure 36 illustrates the trust transfer procedure for a FATP R0-to-R1 handoff.

When the STA detects a degradation of SNR below a threshold, it performs a selective active scan (which will be described in Chapter 7) to find a new candidate AP for handoff. The STA then initiates the FATP handoff by signalling its current associated AP with a FATP request message, which will contain the following information:

- BSSID of the target AP ($MAC_{R1\text{-}AP}$),
- BSSID of the STA's R0-AP ($MAC_{R0\text{-}AP}$),
- MAC address of the STA (MAC_{STA}),
- a new STA identity token (Ns'),
- a random nonce, SNonce, and
- a validating key encrypted with the shared key.

[18] More detail on how the RADIUS enables the encryption of packets exchanged between APs is available in the IEEE 802.11f standard.

When the current AP receives the FATP request, it decrypts the attached validating key and verifies the authenticity of the request (refer to Section 5.2.2). If the verification is successful, the frame will be accepted, otherwise a FATP-Failure message is returned to reject the request and no state information is stored. This is shown as step one in the figure. Because the signalling request frame is protected, it prevents frame spoofing and/or flooding based attacks, which are actually threatening to most of the existing handoff solutions.

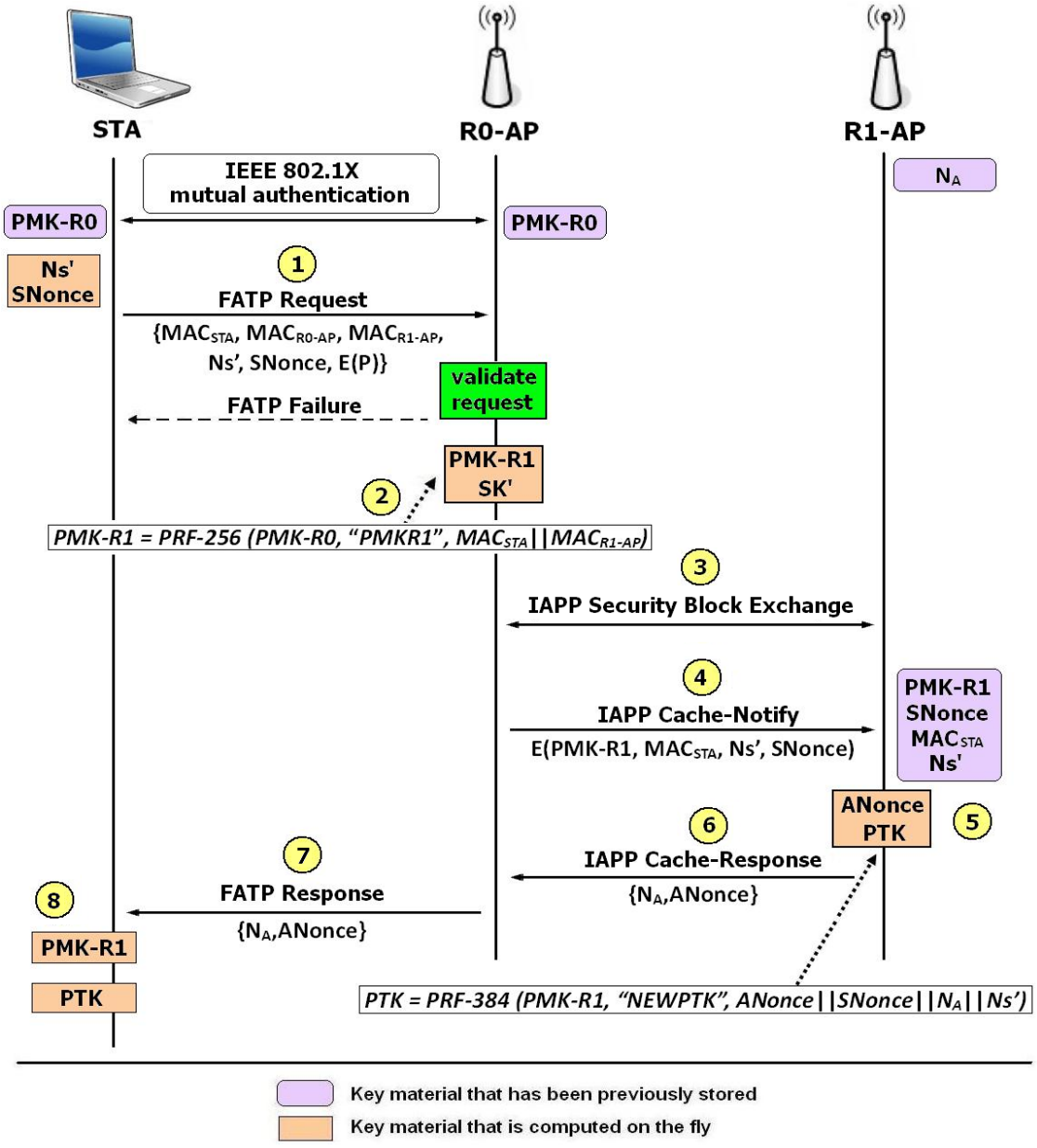

Figure 36: The FATP handoff scheme – Trust transfer for R0 → R1 handoff

In this scenario where the current AP is actually the R0-AP of the STA, a PMK-R1 can be generated locally instantaneously by the current AP using the following equation:

$$\textbf{\textit{PMK-R1}} = \textbf{PRF-256 (}\textbf{\textit{PMK-R0, "PMKR1", }}\textbf{\textit{MAC}}_{\textbf{\textit{STA}}} \,||\, \textbf{\textit{MAC}}_{\textbf{\textit{R1-AP}}}\textbf{)} \quad \ldots\ldots\ldots(1)$$

where the PRF is the pseudo-random function defined in the 802.11i amendment. The function takes the PMK-R0 as the input key as well as parameters including the MAC address of the STA and the new AP. Depending on the security requirements, the shared key (SK) can also be updated using this new PMK with the key update method presented in Section 5.2.3 in page 71. This is shown as step 2 in the figure.

Once the PMK-R1 is ready, it can then be pre-distributed to the target AP that is specified by the STA in the FATP request. To ensure the key transfer is secure, the IAPP security block exchange is performed. This allows the AS to authenticate the new AP to ensure it is genuine and derives encryption keys needed to establish a secure communications between the current AP and the new AP. This is shown as step three in the figure. After that the current AP transfers the security context of the STA, which includes PMK-R1, MAC_{STA}, Ns' and SNonce, to the new AP with an IAPP cache-notify message. The content of this message will be encrypted to ensure confidentiality and integrity. Upon receipt of the cache-notify message, the new AP decrypts it to obtain the PMK-R1 together with other security context information included.

After obtaining the security context attached in the IAPP cache-Notify message and caching the PMK-R1, the new AP now has all the information needed to generate a fresh PTK for the STA. In step five, the new AP produces a random bitstring, ANonce, and computes a PTK using the following equation:

$$\textbf{\textit{PTK}} = \textbf{PRF-384 }\textbf{\textit{(PMK-R1, "NEWPTK", ANonce||SNonce||}}\textbf{\textit{N}}_{\textbf{\textit{A}}}\textbf{\textit{||Ns')}} \quad \ldots\ldots\ldots(2)$$

The PTK is computed based on the PMK-R1, as well as the STA and the AP's random nonces and identity tokens. Including the identity tokens in the PTK computation not only provides randomness to the key (because N_A and N_S' will be different every time a handoff happens) but also allows the STA token to be updated for frame authentication after handing off to the new AP. Even if an attacker can manage to

intercept Ns', N_A and the nonce values, the PTK still cannot be compromised because the attacker has no access to the PMK-R1 generated by the R0-AP on the fly, unless the R0-AP itself or the IAPP security context transfer process is compromised.

In the FATP scheme, each AP maintains a PTK-Cache table, which is a database for storing information about the cached PTKs. The content of the PTK-Cache table includes the STA's MAC address, a PTKID, and the cached PTK. Table 5 shows the table entry.

MAC_{STA}	PTKID	PTK

Table 5: PMK-Cache Table Entry

The PTKID is a bitstring that can uniquely identify the cached PTK. The computation of the PTKID uses the following equation:

$$PTKID = \text{Truncate-128 (SHA-256 (PTK} \,||\, Ns' \,||\, N_A \,||\, MAC_{STA} \,||\, MAC_{R1\text{-}AP})) \quad \ldots.(3)$$

Each PTK-Cache table entry has a pre-defined timeout period. If the new AP does not receive a re-association request from the STA after the timeout period, the entry will expire and the cached PTK will be deleted. The entry is also removed after the STA has successfully re-associated with the new AP.

Once the PTK computation and caching is complete on the new AP, an IAPP Cache-Response message is replied to the current AP to indicate that the key caching is successful, as shown in step six. This response message is extended to also include the new AP's identity token and ANonce.

In step seven, the current AP confirms to the STA that the FATP trust transfer process to the specified new AP has been successful. It does this by replying with the FATP response message to the STA. The new AP's identity token and ANonce are attached to allow the STA to generate the PTK. This message also serves a similar purpose as the first message of the four-way handshake, which is to deliver the nonce value. However, the difference is that this frame contains a MIC that is computed over the whole message using the new PTK.

Finally in step eight, the STA computes the PMK-R1 and the PTK using equations (1) and (2), respectively. The MIC is also verified with the computed PTK. If the MIC is successful, the PTK will be cached and used later when the STA re-associates to the new AP. The benefit of including this MIC is twofold: it allows the STA to ensure that the new AP is actually holding an identical PTK and the MIC also guarantees that the key material carried in FATP response frame has not been tampered with.

6.3.3.4 FATP Trust Transfer Procedure: R1-to-R1 handoff

Figure 37 illustrates the trust transfer procedure for a FATP R1-to-R1 handoff.

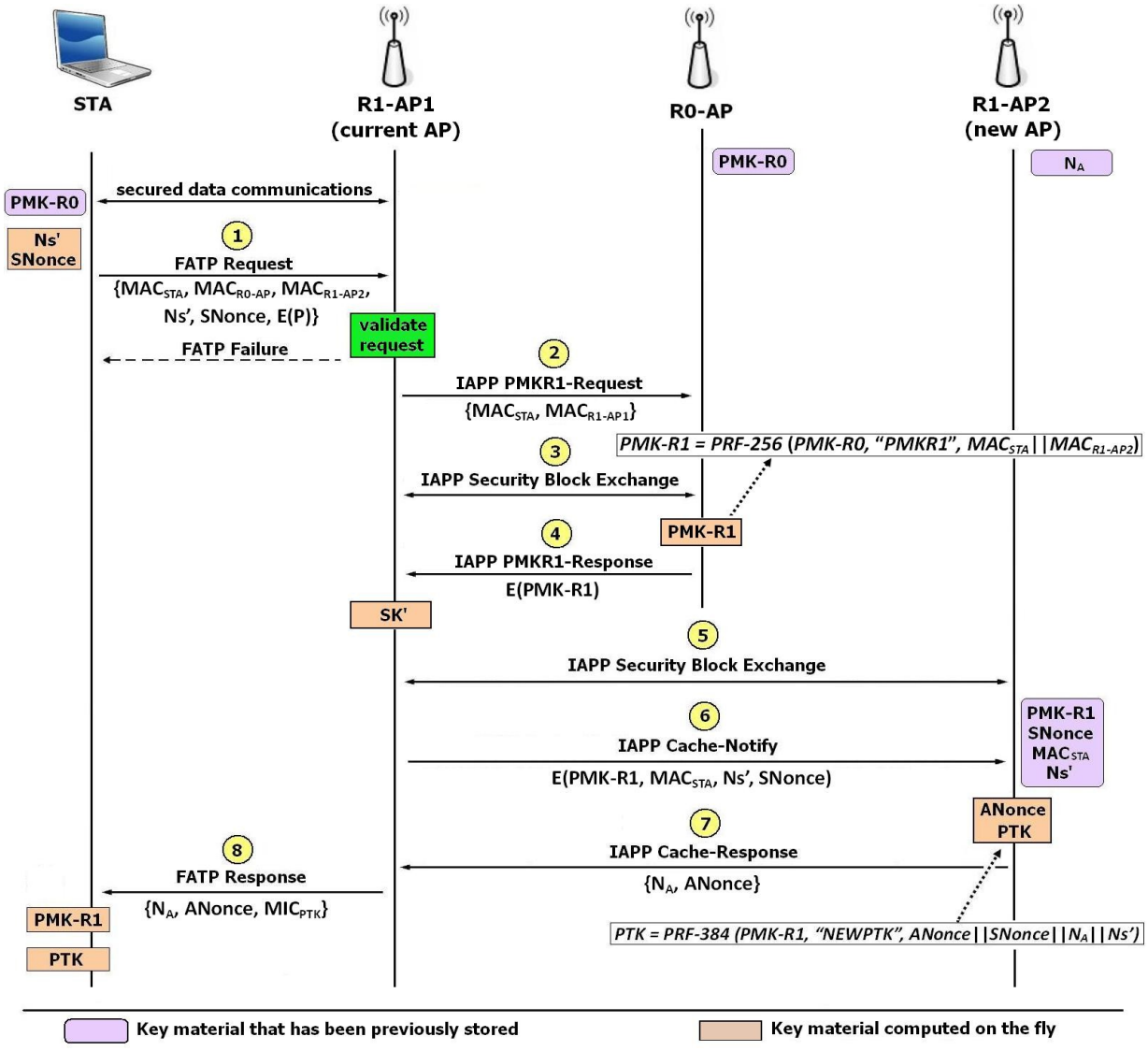

Figure 37: The FATP handoff scheme – Trust transfer for R1 → R1 handoff

In this scenario, the STA is roaming from AP1 to AP2 which both are R1-APs. It is the same as for the R0→R1 handoff scenario where the STA initiates the FATP trust transfer by sending the FATP request to the R1-AP1 which can then validate the request using the attached validating key. This is shown as step one in the figure. From the request, R1-AP1 will be informed with the MAC address of the STA's R0-AP and the desired target new AP (R1-AP2). Because R1-AP1 does not know the PMK-R0 of the STA, it has to request for a PMK-R1 to be generated and transferred from the R0-AP.

To request for a PMK-R1, R1-AP1 sends an IAPP PMKR1-Request to the STA's R0-AP. To establish a secure channel for delivering the key, R0-AP verifies R1-AP1's identity through RADIUS and then security blocks are exchanged. After mutually establishing encryption keys, R0-AP generates the PMK-R1 using Equation (1) and sends it to R1-AP1 with an IAPP PMKR1-Response message, which is encrypted. Upon receipt of the response, R1-AP1 decrypts the message to obtain the PMK-R1, and, if required, updates the shared key using this PMK-R1. This process is depicted as steps two to four in the figure.

Once R1-AP1 obtains the PMK-R1 from R0-AP, it securely transfers the key to the new AP, R1-AP2, using the same IAPP caching method as previously described in the R0→R1 handoff scenario. This process is depicted as steps five to seven in the figure. After receiving the cache notify request, R1-AP2 computes the PTK using equation (2) and the corresponding PTKID using equation (3), and caches the information in the PTK-cache table.

Finally, the identity token of R1-AP2 and ANonce are sent to the STA in the FATP response message, which also contains the MIC computed over the message using the new PTK. After receiving the FATP response, the STA computes the PMK-R1 and the PTK. The MIC is then verified to ensure that R1-AP2 has actually cached an identical PTK.

6.3.3.5 Fast Re-association

When the STA has confirmed that the FATP trust transfer process has been completed and the target AP has cached an identical PTK, it is ready to initiate a re-association process with the new AP to change its point of attachment in the WLAN. The fast re-association procedure is depicted in Figure 38 below.

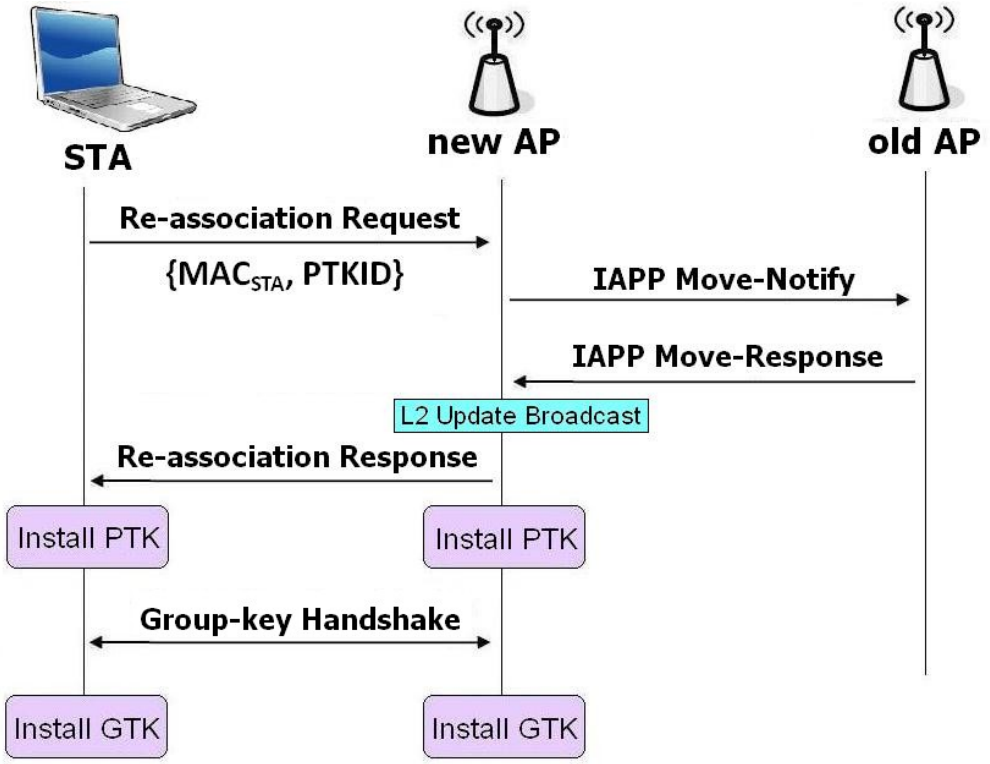

Figure 38: The FATP Fast Re-association Procedure

The STA initiates the re-association process by sending a re-association frame to the new AP. In order to verify the trust relationship with the new AP, both the STA and the new AP must ensure that they both possess a legitimate copy of the PTK. To confirm this, the STA computes the PTKID using Equation (3) and attaches it to the re-association request frame. Upon receipt of the request, the new AP will look up the PTK-cache table for the entry that matches the STA's MAC address and compare the given PTKID from the STA against the PTKID previously cached. If a match is found, both parties are considered legitimate with respect to the trust relationship transferred from the STA's R0-AP.

The new AP then sends the IAPP MOVE-Notify message to the old AP to notify the STA's new association. The old AP deletes the association information for this STA and replies with the MOVE-Response. The new AP then updates its association table and stores the STA's association information.

When a STA moves from one AP to another AP, the link-layer routing state in switches need to be modified accordingly, otherwise packets destined to the STA may be incorrectly routed to the old AP. To do this, the new AP broadcasts a Layer 2 Update frame, on behalf of the STA, to inform all Layer 2 devices (e.g., switches) in the DS to update their forwarding table for this STA. Finally, the new AP responds to the supplicant with the re-association response to indicate that the re-association is successful and can start using the cached PTK. To complete the handoff, only an 802.11i Group-Key Handshake is needed to establish GTK for securing multicast and broadcast traffic.

6.4 Security Analysis

The FATP handoff scheme shortens the long handoff latency by pre-initiating a trust transfer process prior to handoff and performing a lightweight, fast re-association process to verify the transferred trust relationship after moving to the new AP. The security of the transferred trust relationship with the new AP is based on the PMK-R1, which is derived from the PMK-R0 generated by an initial IEEE 802.1X authentication. Because the PMK-R0 is a result of a successful 802.1X authentication, it is considered secure and represents the trust authority. By generating a PMK-R1from the PMK-R0, it provides a means to allow the trust authority to be delegated to the possessor of the PMK-R1, from the R0-AP who owns the PMK-R0. This concept is the same as the IEEE 802.11r key hierarchy and allows the re-establishment of a trust relationship with a new AP to be based on existing trust relationships derived from full 802.1X authentication, without needing the AS to perform any handoff related re-authentication work.

Based on the FATP handoff scheme, the following benefits are achieved:

- **Reduced AS load and handoff traffic overhead without breaking existing 802.11i security:** the AS does not need to participate in the re-authentication for roaming STAs. The AS is only required to assist in the IAPP security context transfer, which only requires a two-message exchange (Access-Request and Access-Accept). The traffic overhead in the DS and the AS resources are therefore reduced. This is a substantial network performance improvement, particularly in an enterprise WLAN where handoff frequency is high.

- **Protected authenticated STAs from re-associating to malicious APs:** requesting a FATP trust transfer to a new AP involves establishing a secure channel with the new AP using IAPP. Therefore, each legitimate AP must have a valid shared secret with the AS and is required to register with the AS when it first joins the network. A fake AP will fail to be authenticated by the AS and will not be able to proceed to the IAPP secure channel establishment. Thus, STAs are protected from rogue AP attacks and Man-in-the-Middle attacks.

- **Shorter handoff latency and outage time:** as the STA initiates a FATP trust transfer, it can still carry on the ongoing communications without interruptions, while the trust transfer process is taking place in the wired network. Therefore, the FATP trust transfer does not contribute to the communications outage and no packet loss is anticipated. The fast re-association does not require a four-way handshake to generate a PTK because it is already computed and cached prior to handoff. Each STA-AP association will always bind to a fresh PTK derived as with the 802.11i four-way handshake.

- **No unnecessary sharing of PMKs among APs:** the FATP handoff scheme allows the STA to decide which candidate APs to handoff to. The STA can also repeat the FATP trust transfer request to multiple APs to pre-cache PTKs. Thus, there is no unnecessary sharing of PMKs among APs in the ESS unlike most of the PKD based handoff solutions.

In the FATP handoff scheme, PMK-R1s and PTKs are derived from the initial PMK-R0s, which are produced after an 802.1X authentication. To allow the PMK-R0s to be

refreshed, the PMK-R0s maintained by a R0-AP are only valid for a configurable lifetime. A R0-AP will keep a timer for each PMK-R0 stored. When a R0-AP receives a PMKR1 generation request from a R1-AP while the associated PMK-R0 is expired, a response indicating expired PMK-R1 will be returned to the requesting R1-AP. The R1-AP will respond with an ACK so the R0-AP can safely remove the PMK-R0 for this STA. The R1-AP will then need to perform a full 802.1X authentication with the STA and become the R0-AP responsible for the generation and transfer of PMK-R1s for this STA upon request until this PMK-R0 is expired.

6.5 Chapter Summary

The original 802.11i pre-authentication scheme is not scalable because it requires every AP to remember PMKSAs for all STAs that could possibly roam to it but have not yet. Each STA has to authenticate to each AP, instead of the infrastructure as a whole, and therefore a complete EAP-TLS authentication to produce the required security keys (e.g., PMK, PTK, and GTK) is still required. This is not efficient for handoff and is a time consuming process. Furthermore, STAs have to also guess which AP they may hand off to and therefore how many APs to pre-authenticate with. To provide a better handoff experience, it is critical that before the STA moves to the new AP, some form of pre-computed security keys are already shared between the new AP and the STA. This can reduce the re-authentication delay considerably and is the main goal of the proposed FATP handoff scheme introduced in this chapter.

To pre-compute security keys without involving the AS, the three tier 802.11r key hierarchy is used in the proposed solution. The key idea behind 802.11r to reducing the re-authentication latency is to delegate the authentication authority to the R0KH and therefore, eliminate the long latency back to the AS. Based on the same principle, the proposed FATP scheme allows a STA to establish security state at the target AP prior to handoff with the concept of trust transfer (i.e., delegating trust authority by delivering PMK-R1s derived from the initial PMKs to the new AP). A fast re-association process confirms the transferred trust relationship, avoiding delays in

connecting to the DS after AP transition. These techniques allow the FATP scheme to support seamless intra-domain handoffs between APs in the ESS. The security of transferring PMK-R1s in the proposed scheme is based the 802.11f standard, which exploits the strong secret shared between the server and APs.

The FATP handoff scheme does not introduce new security vulnerabilities beyond the current IEEE 802.11i standard, and the implementation can preserve the behaviour of legacy STAs and APs. However, it requires two new 802.11 management frames (FATP-Request and FATP-Response) and two additional IAPP messages (PMKR1-Request and PMKR1-Response) to be introduced.

This chapter has covered background information of IEEE 802.11 WLAN handoffs and a brief analysis of existing handoff solutions. The proposed FATP handoff scheme is described in detail, including the handoff trust model and the context transfer. The experimental results of the FATP handoff scheme will be presented and discussed in Chapter 8.

Chapter 7

Location Management based Scanning for Enterprise WLANs

Running real-time multimedia services such as VoIP over WLAN requires seamless and continuous network connectivity. Typically, the inter-packet delay for such applications should be kept within 50ms so that there is no noticeable disruption to the voice/video quality. The VoIP session could be dropped or severe quality degradation is experienced if the STA does not associate with a new AP within 150ms as it moves out from the vicinity of the currently associated AP. Therefore, reducing handoff latency to an acceptable level (ideally less than 50ms [73]) is the key to the successful provision of wireless real-time multimedia services.

A link-layer handoff consists of four main phases: (1) scanning to discover new APs, (2) authentication with a selected AP, (3) re-association with the selected AP, and (4) finally layer-2 update in the DS to deliver packets to the selected new AP. The order of those activities may differ in different handoff solutions. For example, Figure 27 on page 77 shows the handoff process in an 802.11i secured WLAN.

Among those activities, scanning is the most time consuming process which typically incurs a delay of 350 − 1200ms [16, 71]. This accounts to more than 90% of the overall handoff latency. Although the re-authentication and re-association latency could be reduced with the proposed FATP scheme, seamless roaming is still not possible without also reducing the scanning delay. The FATP handoff scheme also requires the STA to select a target AP and provide its BSSID in order to perform trust transfer prior to handoff.

In this chapter, a new type of selective scanning (i.e., without needing to scan all available 11 channels) solution based on a location management scheme is proposed. The location management scheme is a client-server process that provides to the roaming STA the topology information about neighbouring APs, so the number of channels to scan is reduced to a small subset. To further reduce the scanning latency, the combination of two existing techniques, namely background scanning in power saving mode and IP-based probe response, are used to eliminate the waiting times spent in scanning each channel. The performance evaluation shows that the proposed solution can result in a 90% reduction in the handoff latency from standard handoff procedure.

This chapter first provides the background information about the current scanning methods and discusses some related works. The proposed Location Management based Selective Scanning (LMSS) will be described in Section 7.3. The implementation details and performance results will be covered later in Chapter 8.

7.1 Background

The IEEE 802.11b/g standards have specified 11 operational channels in the 2.4 GHz frequency band. Out of these 11 channels, only three channels 1, 6 and 11 do not overlap, as shown in Figure 39. So, in a well-configured WLAN, most of the APs operate on those non-overlapping channels. Two adjacent APs normally would not use the same channel frequency in order to avoid co-channel interference.

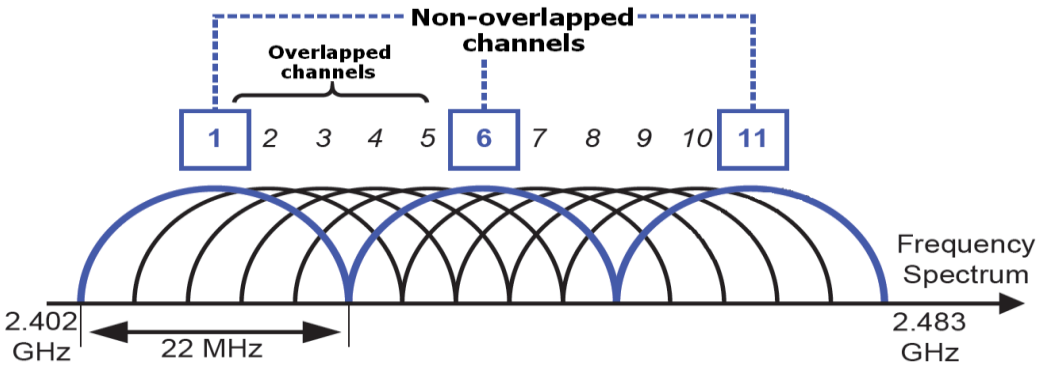

Figure 39: IEEE 802.11b/g frequency spectrum

To maximise spectrum efficiency and user capacity, a WLAN is typically deployed with the cellular concept, depicted in Figure 40. Each hexagon cell represents the radio coverage of an AP. Having multiple overlapping cells can provide network access to cover a larger area. The adjacent APs will operate at different frequencies (ideally non-overlapping frequency channels) to avoid interference; non-adjacent APs can still operate on the same channel because the distance reduces the signal strength, and thus minimises the interference. The arrow in the figure represents the maximum distance between cells using the same channel. With this cellular structure, only three channels are required to cover a large area.

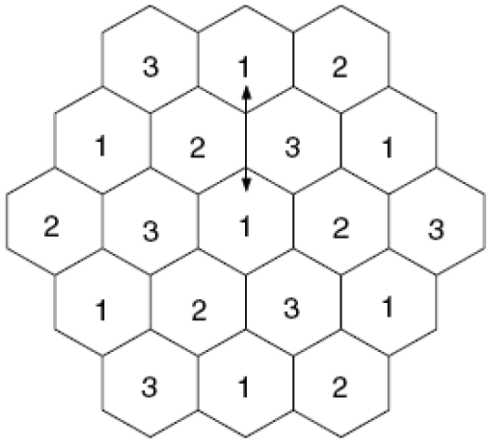

Figure 40: Typical WLAN deployment with three-channel cellular structure

As APs send beacons at predefined intervals (e.g., 100ms), it allows STAs to monitor the signal-to-noise ratio (SNR) of the link to their AP based on these received beacons. As a STA moves from one AP to another, the SNR measured from the current AP will decrease. When the SNR gets below a pre-defined threshold value, the initiation of a handoff procedure will be triggered. Before the connectivity to the network is lost, the STA needs to find potential new APs to re-associate by scanning (or probing process). This is exactly the purpose of the discovery phase of a handoff process. In order to allow a STA to find nearby APs, IEEE 802.11 defines two scanning approaches: the passive scan and the active scan.

In passive scanning, the STA switches to a candidate channel and listens for beacons from nearby APs which announce their presence. This is repeated for all the 11 channels in sequence. Because the STA has to spend at least a beacon interval time in

each channel, passive scanning usually incurs significant delay. In the worst case, passive scanning can take at least a second to scan the other 10 channels (10*100ms).

To reduce this delay, most 802.11 implementations use active scanning which actively broadcasts a probe request on each channel to force an access point to respond immediately.

The detailed active scanning procedure is as follows:

1. The normal channel access procedure (i.e., CSMA/CA) is performed to gain control of the wireless medium.
2. The STA sends a probe request containing the broadcast address as its destination.
3. A probe timer is started.
4. The STA waits for probe responses.
5. If no response has been received by *MinChannelTime*, the STA switches to the next channel and repeats the above steps.
6. If one or more responses are received by *MinChannelTime*, the STA waits until *MaxChannelTime* and processes all of the responses received by this time.
7. The above steps are repeated for the next channel.

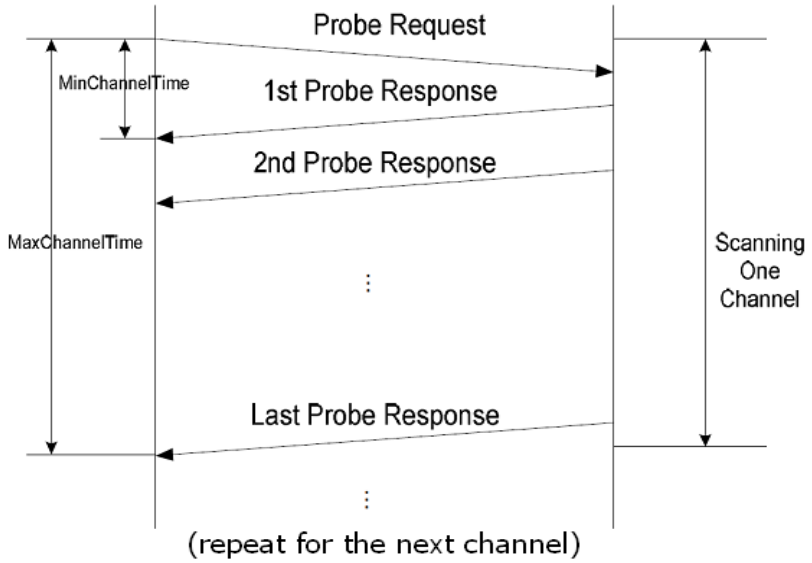

Figure 41: IEEE 802.11 Active Scanning

The IEEE 802.11 standard does not specify values for *MinChannelTime* and *MaxChannelTime*, so they are device-dependent. The active scanning delay bound can be expressed as

$$N \times MinChannelTime \leq DELAY \leq N \times MaxChannelTime$$

where N is the number of channels available.

In most implementations, the MaxChannelTime is set to 50ms, which implies that the scanning delay in the worst case is at least 500ms. This is well beyond the tolerable limit for VoIP and multimedia applications. Therefore, reducing the scanning latency has been an important issue needing to be addressed in order to achieve seamless handoffs.

7.2 Related Work

Many research studies have proposed solutions to reduce the scanning latency in IEEE 802.11 networks. Some solutions such as [67] attempt to completely eliminate the scanning delay by using multiple radios. For example, if a STA has multiple radios (i.e., equipped with more than one NICs), it can use one radio for data communications and another for scanning and pre-authentication with candidate APs at the same time. Although this approach avoids the disconnection of the ongoing data sessions, from the view point of users, it is not cost effective to install additional NICs.

Some research efforts such as [82] leveraged the active scanning mode and derived optimal values for *MinChannelTime* and *MaxChannelTime* from their measurement results and analytical models. Specifically, they used more aggressive values of 1ms and 10ms for *MinChannelTime* and *MaxChannelTime*, respectively. By using these reduced timer values, the channel probe delay is reduced. However, all channels still need to be scanned.

In order to reduce the number of channels to scan, solutions based on Neighbour Graphs (NG) [75, 83] are proposed. Using the NG, the set of channels on which neighbouring APs are currently operating and the set of neighbouring APs on each

channel can be learned by the STA. Based on this information, the STA can determine whether a channel needs to be probed or not. The NG approach based solutions require extra functionality to be implemented at both STAs and APs that infer the WLAN topology. Their results show that the probing latency could be reduced to around 30-40ms. However, these NG schemes have the following drawbacks. Firstly, both STAs and APs require major software changes and modifications to the existing 802.11 state machine. Secondly, considerable signalling overhead between an AS and APs is involved for the creation and maintenance of the network topology.

Another class of solutions being proposed is to reduce the time spent on scanning each channel. A typical solution of this class is the SyncScan [84]. SyncScan is a technique for continuously tracking nearby APs by passive probing in the background so that the scanning latency can be eliminated. This scheme requires that the clocks on all the APs are synchronised and each AP broadcasts its beacon based on a well-defined time schedule. By switching to channels on a scheduled basis and looking for beacons, the client can learn about the neighbouring APs so that no extra scanning is needed at the time of handoff. Although SyncScan greatly reduces the scan time to only a few milliseconds, a very precise time synchronization of all parties is needed, which may not be easy and practical to do. There is also a hidden cost since a STA has to periodically postpone its current connection to listen to beacons of other APs. Therefore, the STA could suffer from possible loss of frames during channel switching and performance degradation in congested networks.

Another scanning solution called Proactive Scan [85] uses triggers to proactively scan the channels to find the best AP. This technique decouples the time-consuming channel scanning from the actual handoff and eliminates channel scan delay by performing scans prior to handoff and interleaving it with ongoing traffic in a non-intrusive way. However, this approach does not consider the impact of this periodic scanning on the energy consumption on mobile STAs, which are normally limited by battery lifetime.

A similar solution for 802.11 mesh networks called Opportunistic Scanning [86] also attempts to perform scanning prior to the actual handoff to avoid handoff delay. The main difference is that the Opportunistic Scanning proposes the use of the 802.11

power saving (PS) mechanism as a signalling protocol to pause the ongoing communications while the STA switches to other channels for scanning. This allows the current AP to buffer the packets destined to the STA while the STA is performing scanning in the "sleeping mode". When STA returns back to the normal operation mode, the AP will send the buffered packets to the STA so that there will be no packet loss. However, this solution still requires scanning through all channels. Based on the same idea of using 802.11 power save (PS) mode to buffer packets while STA is performing scanning, a background scanning solution is proposed in [87] which further uses a location server to help reduce the number of channels required to be scanned. Although those solutions can reduce handoff latency by performing scanning in power saving mode prior to handoff, disruption to the ongoing communications is still possible if the number of channels to scan is not small or there are many neighbouring APs.

A very different type of solution for reducing the channel waiting time called Fast-scan, is proposed in [88] for IEEE 802.11-based fixed relay radio access networks and ad-hoc networks. The idea of this solution is that instead of sending probe responses directly as MAC-layer unicasts, the AP replies with an IP-based probe response to the scanning STA. To perform this fast-scan, the STA first broadcasts an extended probe request containing its IP address on one channel and changes to the next channel directly thereafter. After sending probe requests on all the channels, the scanning STA returns to its original channel to continue receiving and sending normal data packets while awaiting probe responses. The APs upon receipt of the extended probe requests will reply with a UDP-based probe response to the IP address of the scanning STA. With this approach, the STA only has to leave its current channel for sending probe requests, and directly thereafter it can return to the original channel. The benefit of this cross-layer solution is that scanning STAs can continue to send and receive normal data packets while receiving probe responses. Therefore, the latencies due to response waiting times are eliminated. Their result shows that the total scanning latency can be reduced to around 80ms with the fast-scan solution.

7.3 Proposed Fast Scanning Scheme

Once a handoff decision has been made, the STA must scan each channel to lookup potential new APs in its radio range. Time could be wasted if the STA scans the channels that are possibly empty. It is a common practice that a well-configured WLAN is deployed with two, or even three, overlapping BSSs at the radio signal's boundaries. This is because having too many overlapping APs in one area will result in poor radio quality due to interference from the neighbours. Considering this fact, it would be unnecessary to scan all the channels to find candidate APs for handoff. Therefore, reducing the number of channels to scan the minimal required set is one of the important aspects of reducing the handoff latency. To support this feature without making major changes to the existing 802.11 protocol, the thesis proposes a Location Management-based Selective Scan (LM-SS) scheme to help STAs to only scan the required channels. The decisions are based on the topology information provided by a Location Server. The Location Server is a network entity that manages and stores the AP's topology at the network side and a STA can request this topology information when required. Unlike neighbour graph based solutions, the implementation of the location management is very simple but effective and the STAs are not required to run any prediction algorithms or perform any resource-expensive computations.

The initial experimental results obtained demonstrate that the handoff latency is drastically reduced as a result of the reduced number of channels to scan with the location management scheme. However, if a STA is required to scan more than 3 channels, the probing delay is still significant (more than 100ms) because for each channel the STA has to wait until the *MaxChannelTime* expires. To further reduce the probing delay and waiting time in each channel, the proposed scanning scheme uses the IP-based probe response concept introduced in [88] and implements it for 802.11 infrastructure mode based WLANs. The STA leaves the current channel only for the time required to send probe requests and then immediately returns to the channel to resume data communications while still being able to receive probe responses. This way the probing delay is reduced by avoiding the wait for probe responses on the channels to be scanned. To prevent possible packet loss while the STA is out of the

current channel for sending probe requests, the 802.11 power saving (PS) mode is used to signal the AP to buffer packets.

In summary, the proposed scanning scheme, LM-SS, utilises the following three techniques for reducing total handoff latency: selective scanning enabled by a location management mechanism, IP-based probe response and power saving mode provided by the existing 802.11 standard. Each of these techniques will be described in detail in the following sections.

7.3.1 Location Management

As mentioned earlier, the main factor contributing to handoff latency is the scanning delay. If a STA knows exactly the presence of its nearby APs, it can perform selective scanning with unicast probe requests to avoid scanning all channels. Not only could the scanning delay be reduced as a result, unnecessary bandwidth utilisation is also prevented because only those specified candidate APs are required to respond to the probe requests. To achieve this, a location management server is introduced in the backbone network. The server side database management and the location information exchange will be described in the following sections.

7.3.1.1 Dynamic Topology Database Management

When a STA first joins the network, a normal 802.11 active scan is performed to scan all channels. Based on the received probe responses, the STA associates to the AP of highest Received Signal Strength (RSS). After the STA has successfully connected to the AP, instead of discarding the results obtained from the scan, the STA sends the result and the AP's BSSID to the location server. Those results are referred to as *Scan Reports*. A scan report will contain a list of neighbour entries storing information such as the BSSID, IP address, and channel number of the adjacent APs discovered from the active scan.

The location server builds up and maintains an AP topology database based on scan reports sent from STAs in the WLAN. The AP topology database is a linked list of data structures containing the following information:

- Subject AP (6 bytes): the BSSID of the AP to which STA is currently connected
- Neighbour AP (6 bytes): the BSSID of the an adjacent AP
- Channel (1 byte): the channel number used by a neighbour AP
- IP address (4 bytes): the IP address of a neighbour AP
- Timestamp (4 bytes): current timestamp when an entry is added

This topology database can be represented in a table format as shown in Table 6 below. Each entry of the topology table is indexed by the subject AP (i.e., the STA's current AP).

Subject AP (BSSID)	Neighbour Entry			
	Neighbour AP (BSSID)	Channel Number	IP Address	Timestamp
AP1	AP4	1	192.168…	…
	AP5	6	192.168…	…
	…	…	…	…
AP2	AP1	11	192.168…	…
	AP3	6	192.168…	…
	…	…	…	…
…	…	…	…	…

Table 6: AP topology database maintained by the location server

When a scan report is received, the location server extracts the BSSID of the STA's current AP and tries to locate the corresponding subject AP entry for this AP by indexing through the table. If an entry is found, the neighbour entries attached in the scan report will be checked against the existing records in the database. For identical

neighbour entries, the timestamps are updated; for new neighbour entries, they are added to the database. The location server also periodically checks the timestamp freshness for each neighbour entry in the database. If an entry is too old (e.g., greater than 3 days), it indicates that this AP has not been found over this period, possibly no longer exists, and the entry will be automatically deleted. This mechanism allows the location server to dynamically update the AP topology database. As the network operation time progresses, the topology information stored in the database will become more and more accurate and complete. To speed up the learning process, the database can also be manually edited to specify neighbour relationships.

7.3.1.2 Location Information Exchange

Figure 42 below illustrates the proposed location management mechanism. Although the focus of this research is on intra-domain handoffs, this scheme can also be applied to manage inter-domain (i.e., across different subnets) handoffs, as shown in the figure.

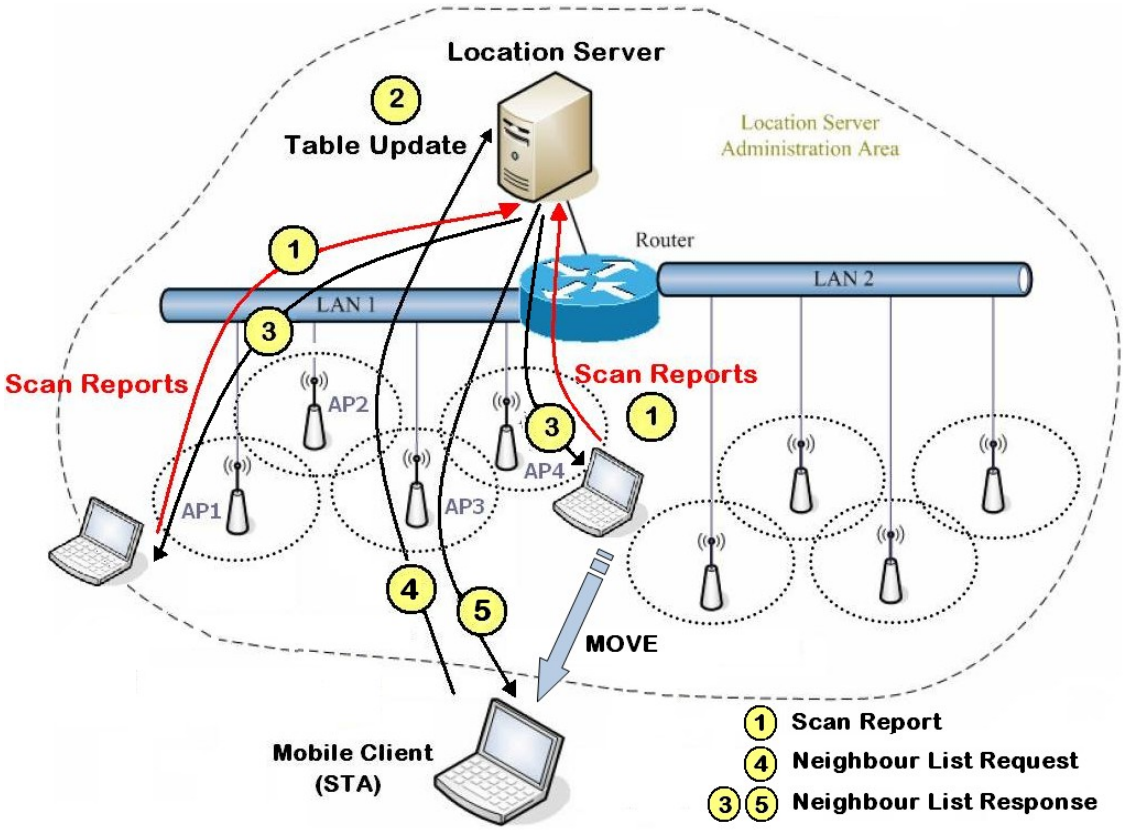

Figure 42: Proposed Location Management Scheme and Network Structure

The proposed location management scheme defines the following three messages. More details regarding the packet format and header fields will be presented later in Chapter 8.

- **Scan Report**: contains a list of neighbour entries
- **Neighbour List Request**: used by a STA to request neighbour information from the location server
- **Neighbour List Response**: the reply to the above request to inform the STA about neighbour APs that are adjacent to the STA's current AP

As illustrated as step 1 in the figure, when STAs initially connect to the network (e.g., via AP1 and AP4) they will automatically connect to the location server and upload their scan reports. After processing the scan report and updating the topology database (step 2), the location server replies with a neighbour list response to inform them about the neighbouring APs in their current area, as shown as step 3. Based on this information, a STA will only need to perform selective scanning with unicast probe requests to the given list of neighbour APs when a handoff is required.

After a successful re-association with a new AP (AP3 in the figure) as a STA moves, the STA needs to request the location server for the neighbour information with respect to its new AP for the next handoff. To do this, the STA sends a neighbour list request, which contains the BSSID of its current AP, to the location server and the server will return a list of neighbour APs in the STA's current location with a neighbour list response message. This process is depicted as step 4 and 5 in the figure. As AP3 only has two neighbours, namely AP2 and AP4, the STA will only need to scan the channels used by those APs when handoff is needed. Based on the received probe responses, the STA will re-associate with the AP that has the highest RSSI and repeat step 4 and 5 to prepare for the next handoff.

For this proof-of-concept implementation, the IP address of the location server is manually configured on each STA in the testbed network. The STAs are running a user-space application which is responsible for handling the neighbour information

request and exchange with the location server. The location management application running on the location server is summarised in the protocol flow diagram below.

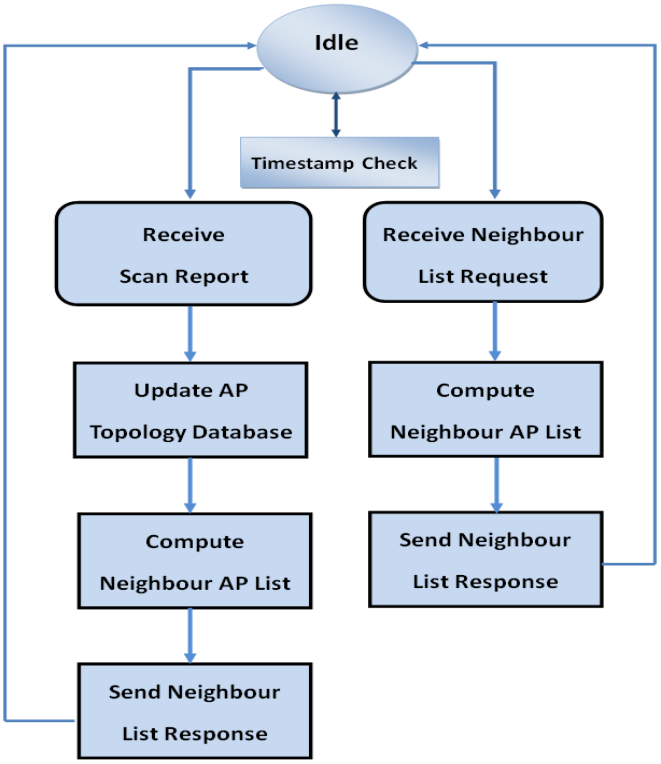

Figure 43: Flow Diagram of Location Management for Server Application

7.3.2 IP-Based Probe Response

The STA continuously monitors the received signal strength (RSS) based on the beacons received from the associated AP. When the RSS is below a pre-defined threshold value, a handoff process is initiated. With the given list of neighbour APs from the location server, the STA will need to switch to each of the channels used by the neighbour APs and send a unicast probe request in order to obtain the signal strength indication from those APs using the replied probe responses. To eliminate the probe response waiting time in each channel being scanned, the IP-based probe response concept from the fast-scan [88] is adopted in the proposed LM-SS scheme.

The scanning STA loops through the cached neighbour entries and, for each neighbour AP, switches to the corresponding channel to send an extended probe request. An

extended probe request is a unicast probe request with an additional information element (IE) containing the IP address of the scanning STA. In this thesis, these extended probe requests are referred to as “Fast Probing Requests”.

After sending these fast probing requests to all the neighbours, the scanning STA returns to its original channel and resumes the data communications while waiting for IP-based probe responses. Those IP-based probe responses are referred to as “Fast Probing Response” in the thesis.

With the fast-scan solution, the probe response waiting phase is governed by the parameters *minReplyWait* and *maxReplyWait*. If no reply has been received within *minReplyWait*, the STA concludes that no neighbours were found; if a reply has been received within *minReplyWait*, the algorithm continues to listen for replies until *maxReplyWait* [88].

One major difference between the fast-scan solution and the LM-SS scheme proposed in the thesis is that this waiting time is not required with the LM-SS scheme. This is because the scanning STA already has the knowledge about the nearby APs and will not expect to receive other responses. Also, because the probe requests sent by the STA are unicasts, no APs other than the request recipients would actually reply. Therefore, there is no waiting time needed after the STA returns to the original channel after probing.

On the AP side, when an AP receives a probe request, it checks whether the request contains a fast probing information element (FPIE). If it does, the AP extracts the scanning STA’s IP address from the FPIE and sends a fast probing response to the extracted IP address. The fast probing responses are UDP packets, which have exactly the same information and format as IEEE 802.11 probe responses [1], but further extended by a field containing the signal to noise ratio (SNR), with which the AP has received the corresponding fast probing request from the scanning STA. The SNR field functions as a link quality feedback to the scanning STA to help it select the new target AP with the best signal strength. On the other hand, if the received probe requests do

not contain a FPIE, the AP will handle those requests with the default procedure as described in the 802.11 standard.

With reference to Figure 42 as an example, the following table represents the cached list of neighbour entries on the STA after it has re-associated to AP3 and received the neighbour list response from the location server.

Neighbour AP BSSID	Channel Number	IP Address	SNR
AP2	11	192.168…	(unknown)
AP4	1	192.168…	(unknown)

Table 7: Cached neighbour list on the STA before scanning

As the STA moves towards AP2, the RSSI of the current AP will decrease. When it is below the threshold, the LM-SS is triggered and a fast probing request will be sent to AP2 and AP4. After receiving the fast probing responses from AP2 and AP4, the SNR value will be extracted and the neighbour entry list will be sorted according to the SNR. The STA will select the AP in the first entry to be the candidate AP to handoff because it has the highest SNR value. In this example, AP2 is selected.

Neighbour AP BSSID	Channel Number	IP Address	SNR
AP2	11	192.168…	35dB
AP4	1	192.168…	10dB

Table 8: Cached neighbour list on the STA after scanning

The STA then performs the proposed FATP handoff process with the selected AP. If the re-association fails, the next AP in the list will be selected for handoff. If no APs are found in the neighbour list, the complete 802.11 active scanning will be performed.

7.3.3 IEEE 802.11 Power-Saving Mode

To minimize packet loss during the probing activities (i.e., channel switching and transmission of fast probing requests on other channels), the power saving (PS) mode defined in the IEEE 802.11 specification [1] is utilised as a signalling mechanism to

pause the ongoing communications between the STA and its associated AP as well as to allow the AP to buffer the packets while the STA is sending out requests in the PS mode.

According to the standard, the power saving STAs that currently have buffered packets within the AP are identified in a TIM (Traffic Indication Map), which is included as an element within all beacons generated by the AP. STAs operating in PS mode are required to listen to the beacons and determine if there is any data buffered for it by interpreting the TIM. Upon determining that there is data currently buffered in the AP, the STA will send a short PS-Poll frame to the AP to request the buffered data.

With the proposed LM-SS scheme, the 802.11 PS mode is modified slightly. Instead of periodically interpreting the TIM in beacons, STAs in PS mode will switch between channels to send fast probing requests to the neighbour APs. The PS-Poll frame is transmitted to the AP to poll buffered data only after fast probing requests are sent to all neighbour APs.

To change between power management modes (i.e., PS mode and normal mode), a STA will inform the AP through a successful frame exchange initiated by the STA. The Power Management bit in the Frame Control field of the frame sent by the STA in this exchange indicates the power management mode that the STA will adopt upon successful completion of the entire frame exchange [1]. The AP will buffer the data for STAs while they are in PS mode.

7.4 Chapter Summary

The scanning phase poses the biggest hurdle for seamless handoffs in WLANs. Most of the existing 802.11 implementations only attempt to scan when a STA's link quality degrades to a point where connectivity is threatened. Furthermore, with the considerable delay and overhead associated with the 802.11 active scanning, the total handoff latency, during which incoming packets are dropped, is typically well over 500ms to one second, where 90% of the delay comes from the scanning phase. The

result is far beyond what can be tolerated by real-time multimedia applications such as VoIP.

In this chapter, a novel scanning scheme called LM-SS is proposed to reduce the handoff latency incurred due to layer-2 scanning in 802.11 WLANs so that quality requirements of multimedia applications can be met. The proposed LM-SS scheme uses the following three mechanisms for reducing scanning latency in order to achieve fast seamless handoffs:

(1) the number of channels to scan is reduced to a smaller set with the AP topology information provided by a location server that implements the location management mechanism described in Section 7.3.1,

(2) the probe response waiting time is eliminated with a cross-layer solution described in Section 7.3.2 where IP-based probe responses are used to provide link quality feedback to the scanning STAs, and,

(3) packet loss is prevented by performing channel switching and probe request transmissions in the 802.11 power saving mode, in which the AP will be signaled to buffer data for the scanning STA while it is away from the operating channel.

Working in conjunction with the proposed FATP handoff scheme in Chapter 6, most of the operations related to handoff are pre-executed prior to handoff actually occurring, including the selection of the target AP and the transfer of a STA's security context. More importantly, these activities will not cause significant interruption to the ongoing data communications. Therefore, real-time data communications with tight QoS constraints can still be met during handoffs. Further implementation details and experiment results will be presented in Chapter 8.

Chapter 8

Implementation and Performance Evaluation

The previous chapters proposed the building blocks that can be combined to enable a DoS-resistant WLAN supporting fast and seamless handoffs. Chapter 5 proposed the APN authentication that protects the WLAN against DoS attacks using client puzzles. The APN authentication also allows the connecting STA and the serving AP to exchange identity tokens, which will be used to authentication subsequent frames that require protection. Chapter 6 proposed the FATP scheme that allows the network to perform security context transfer from the current AP to a new AP for the roaming STA prior to handoff to achieve fast roaming. The identity tokens exchanged from the initial APN authentication will also protect the frames associated with the FATP handoff, so that the handoff process is also resistant to frame spoofing and DoS attacks. Chapter 7 proposed the location management-based selective scanning (LM-SS) scheme to mitigate the long delays associated with scanning activities which constitute more than 90% of the total handoff latency. Combining the FATP and LM-SS schemes form a complete handoff suite that can meet the strict quality of service requirements for the provision of real-time services like VoIP on WLANs.

These solutions have been combined into a prototype implementation. This chapter summarises the implementation of this prototype and presents the results of a testbed evaluation of these proposed schemes.

8.1 Testbed Environment

The proposed solutions and schemes are implemented on the original testbed described in Section 4.2.1 on page 45. A location server and an extra AP are brought in to test the handoff performance under different roaming scenarios. Figure 44 shows the network components in the extended testbed.

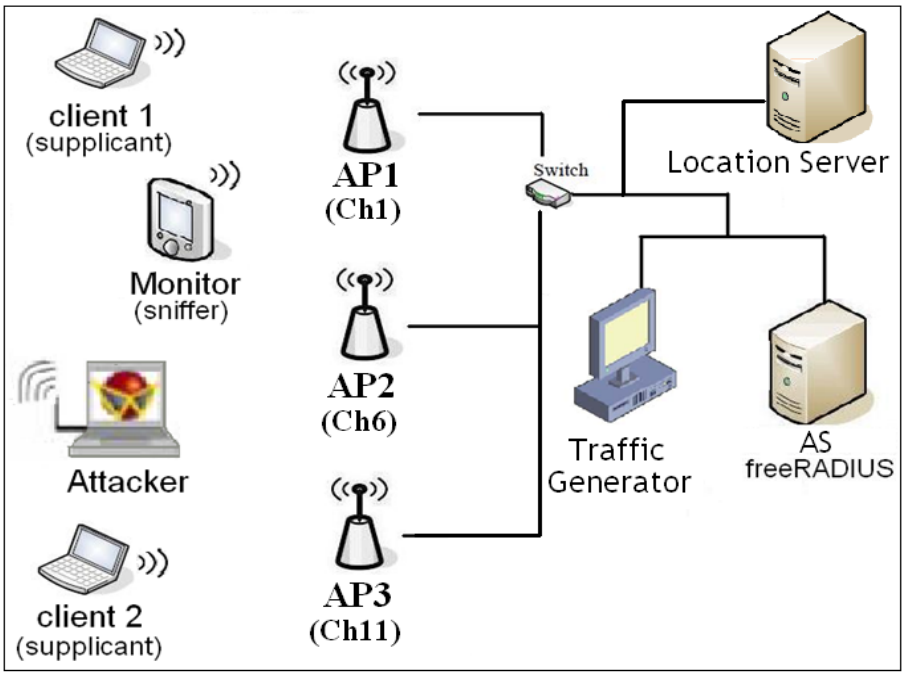

Figure 44: Testbed environment for the evaluation of the proposed schemes

The testbed represents a simplified version of an enterprise WLAN infrastructure and has the following properties that the proposed schemes rely on: the architecture of the WLAN network consists of multiple APs interconnected via a high-speed switched network so that efficient broadcast/multicast communications is available at layer two. In addition, APs are deployed within overlapping radio ranges of each other so continuous wireless coverage is available over the testing areas. A dedicated switch is installed for efficient communications between APs.

In the testbed, all machines are installed with Linux 2.6.25 and run on Pentium 4 2.26GHz CPU with 512MB RAM. NTP service is running on all machines to synchronize their clocks so that accurate traffic latencies can be measured. The APs

and STAs are equipped with an Atheros AR5002G chipset based NIC, which supports IEEE 802.11g (2.4GHz) wireless access and hardware encryption capabilities (e.g., AES) for IEEE 802.11i functionalities. The Madwifi driver (v0.9.4), which is a Linux kernel device driver for Atheros chipsets, is installed on these machines. With the APs, the transmit power of the radio is reduced in order to shrink their cell size to allow handoff to be triggered within a shorter distance from an AP. The STAs and APs are assigned with static IP addresses, as are the location server and the RADIUS server. The wireless clients and the traffic generator STA use TTCP and Iperf tools for traffic generation and throughput measuring. A monitor STA running Wireshark is used to capture the traffic to and from the STAs. The traffic can be later analysed to extract handoff related data and measure the latencies. The testbed is configured to use the strongest upper-layer authentication protocol - EAP-TLS. Figure 45 below shows the packet capture of a complete 802.1X/EAP-TLS authentication procedure.

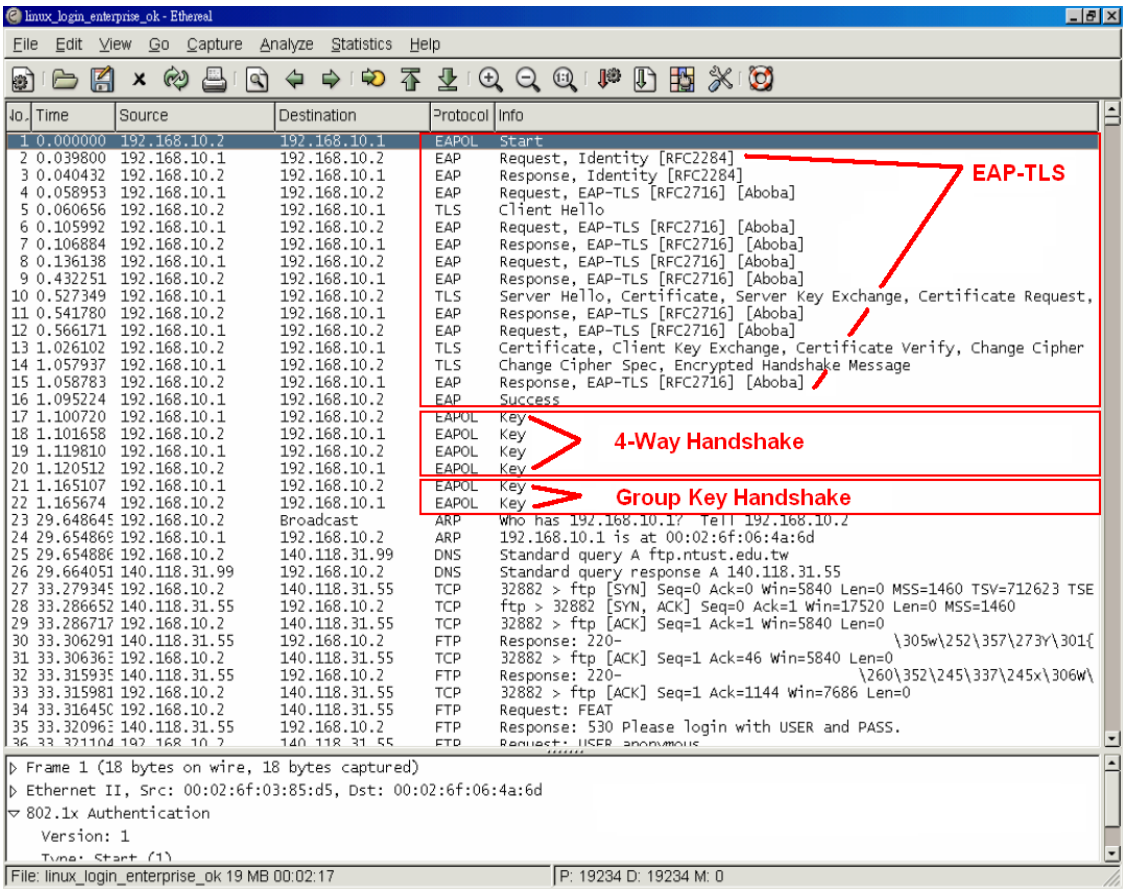

Figure 45: The captured WPA-Enterprise (EAP-TLS) connection procedure

The Madwifi driver consists of four main software modules, namely net80211stack, the Atheros specific "ath" functionalities, Hardware Abstraction Layer (HAL) and rate selection algorithms for managing transmission rates. The implementation of the proposed solutions requires modifications mainly in the net80211 module. A LM-SS user-space application is used to interact with the location server and processes AP topology related information, as well as performing the selective scanning. Figure 46 shows the software architecture of the overall implementation.

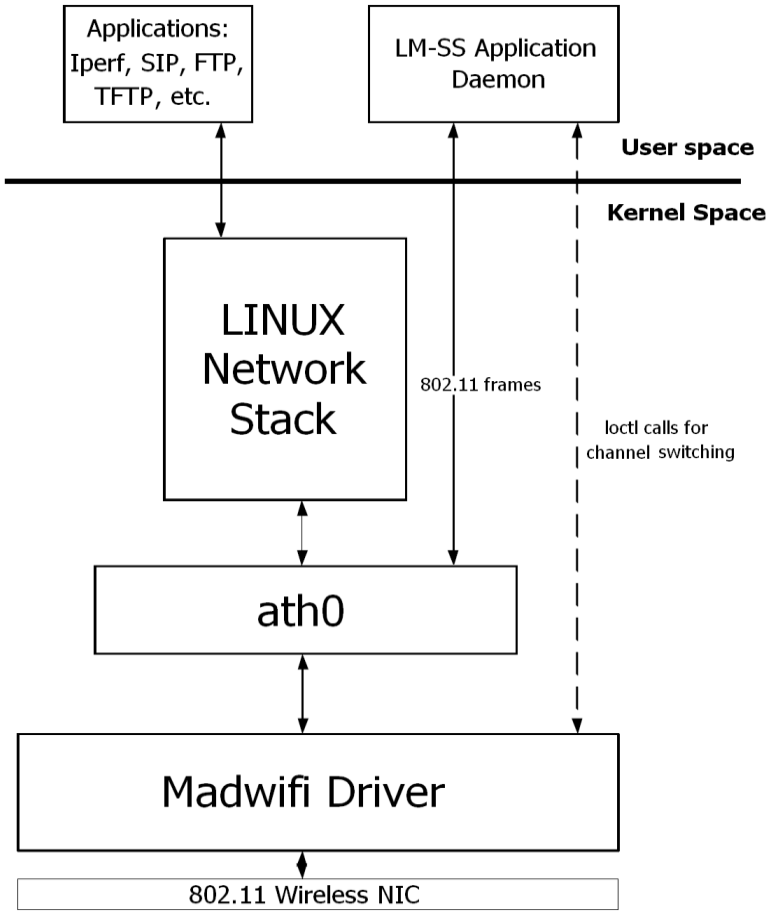

Figure 46: Software architecture of the implementation

8.1.1 APN Authentication Implementation

To support legacy and APN protected STAs within the same ESS, the APN authentication scheme was designed and implemented without changing the existing IEEE 802.11 state machine. All the required information is embedded within the existing frames as Information Elements (IEs).

In order for a client to use APN authentication, the AP indicates that it supports APN authentication in its beacon frames. This also allows an AP to support open system authentication for backward compatibility. In a beacon frame, a two-octet RSN Capabilities field included in the RSN information element is used to advertise an AP's capabilities and supports. Figure 47 below shows the RSN capabilities field format attached in a beacon frame. With the implementation in this thesis, the reserved bit six of the capabilities field is set to indicate the APN authentication support.

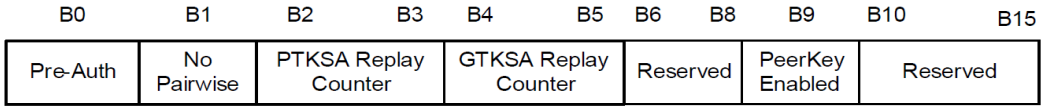

Figure 47: RSN Capabilities field format

The original 802.11 authentication frame is extended to include a new information element, APN-IE, in the frame body, as illustrated in Figure 48 below.

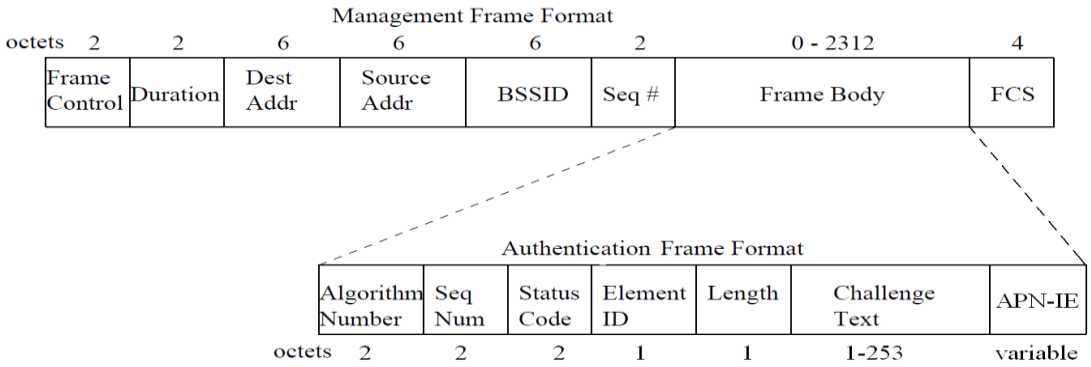

Figure 48: The extended authentication frame format

An information element is defined to have a common format consisting of a 1-octet Element ID field, a 1-octet length field, and a variable-length element-specific

information field, as shown in Figure 49. The Length field specifies the number of octets in the Information field. Each element needs to be assigned a unique Element ID. For the implementation, a reserved value of 17 is used to represent the APN-IE. An APN-IE will contain information such as the timestamp, identity tokens, that will be required during the APN authentication exchange.

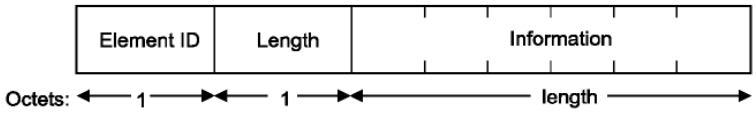

Figure 49: Information element format

The 802.11i amendment specifies that only authentication frames with the authentication algorithm set to Open System authentication may be used within an RSNA. This has been modified to allow both Open System authentication and APN authentication to work with RSNAs. A new authentication algorithm number, 2, is defined to represent the APN authentication. The existing algorithm numbers (0 – Open System, and 1 – Shared Key) and the related functionalities are not changed so that the implementation is backward compatible.

The Challenge Text element in the authentication frame body is originally designed to carry the challenge string used during the exchange of Shared Key authentication. This existing field will also be used by the APN authentication to carry the client puzzle and solution bitstrings without defining additional fields.

In an authentication frame body, the status code field is used to indicate the result of the previous authentication request. The APN authentication uses the following three status codes to indicate different results in an authentication response:

- 0 – Successful
- 15 – Authentication rejected because of challenge failure
- 27 (reserved) – Puzzle challenge required

With reference to Figure 22, which illustrates the APN authentication exchange, the content of the four authentication frames used in the exchange are as summarised in

Table 9 below. The flow diagram in Figure 50 illustrates the APN authentication exchange procedure on the AP.

Message Number	Authentication Algorithm	Seq. Number	Status Code	Challenge Text	APN-IE
1	APN Authentication (2)	1	Reserved	Not present	Client's identity token
2	APN Authentication (2)	2	27	Puzzle	timestamp
3	APN Authentication (2)	1	Reserved	Solution	Client's identity token
4	APN Authentication (2)	2	(0 or 15)	Not present	AP's Identity token

Table 9: Frame contents of the APN authentication exchange

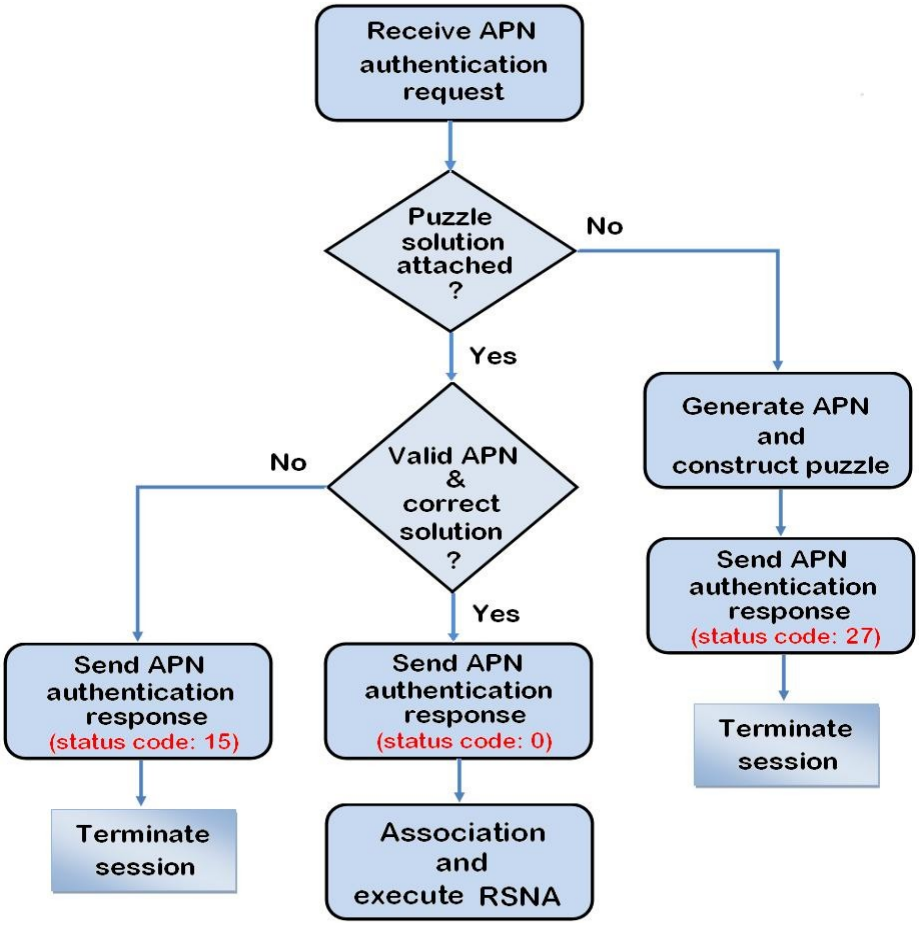

Figure 50: Flow diagram of APN authentication procedure on AP

With the fast re-association implementation, the 802.11 re-association frame is extended to also include the PTKID in the RSN information element in the frame body. For generating identity tokens, a platform-independent C language library called *LibTomMath* is used for manipulating large numbers with the provided APIs. It is modified to compile with Madwifi and to run as a kernel module on a Linux system.

On the client side, some modifications to the wpa_supplicant application were made. Firstly, its control interface was extended to support the APN authentication in additional to the Open System authentication. Secondly, the information associated to each RSNA was extended to also include the identity token of the STA and the AP.

8.1.2 FATP Implementation

The proposed FATP scheme introduced two new management frames: FATP request and FATP response frames. The frame body of a FATP request frame will contain an information element to store the following information: the BSSID of the target new AP, the BSSID of the STA's R0-AP, the STA's new identity token, a nonce, and one encrypted validating key. Similarly, the information element attached in the frame body of a FATP response will contain an AP nonce and the new AP's identity token. A status code field is also included in the frame body of a FATP response to indicate the result of the request. The following values are defined in the prototype implementation:

- Status (0) – successful
- Status (1) – failure
- Status (2) – the specified R0AP is unreachable or does not exist

For the R1→R1 handoff scenario, the current AP does not know the PMK-R0 of the STA, therefore the current AP needs to request the STA's R0-AP to generate a PMK-R0 and transfer it. Two new IAPP messages, PMKR1-Request and PMKR1-Response, are introduced for this purpose. Figure 51 below shows the general IAPP packet format.

IAPP version	Command	Identifier	Length	Data
Octets: 1	1	2	2	0-n

Figure 51: General IAPP packet format

The command field (8-bit integer value) in the IAPP packet identifies the specific functionality (e.g. ADD-notify) of that packet. The original IAPP packets occupy command values from 0 to 6, as shown in Table 10. For the FATP implementation, two additional IAPP commands, PMKR1-Request and PMKR1-Response, are added with command value 7 and 8 as an extension to the original IAPP command set.

Value	Command
0	ADD-notify
1	MOVE-notify
2	MOVE-response
3	Send-Security-Block
4	ACK-Security-Block
5	CACHE-notify
6	CACHE-response
7-255	Reserved

Table 10: IAPP command field values

These new IAPP packets are used by a R1-AP for requesting a PMK-R1 from the STA's R0-AP in the R1→R1 handoff scenario. The data field of a PMKR1-Request packet contains the MAC address of the roaming STA; the data field of a PMKR1-Response packet contains an encrypted PMK-R1, generated by the R0-AP that received the PMKR1-Request packet. These packets are sent using TCP/IP in order to ensure a reliable exchange.

In addition to the introduction of the two new IAPP packets, the existing Cache-Notify packet and Cache-Response packet are also extended to contain additional information to support FATP. The format of the data field of a Cache-Notify packet is shown

below in Figure 52. The MAC Address field will contain the MAC address of the STA that has requested the FATP trust transfer. The Context Length field is a 16-bit integer that indicates the number of octets in the Context Block field. The Context Block is a variable length field that contains the actual security context information to be forwarded to the new AP for the roaming STA indicated by the MAC Address.

Address Length	Reserved	MAC Address	Sequence Number	Current AP	Context Length	Length of Context Block	Context Timeout
Octets: 1	1	n = Address Length	2	n	2	m = Length of Context Block	2

Figure 52: Data field format of a Cache-Notify packet

The context block in the Cache-Notify packet will contain the following information:

- PMK-R1: the PMK to be used to generate PTK for the new handoff session,
- N_S': the STA's new identity token for the new handoff session, and
- SNonce: the STA's random nonce for computing a PTK.

These values are stored in the context block field as a series of information elements of the format as shown in Figure 49.

The Cache-Response packet is sent from the handoff new AP to the STA's current AP that sent the Cache-Notify packet. The data field of an original Cache-Response packet carries the MAC address of the roaming STA and the corresponding sequence number, as shown in Figure 53. This packet is further extended to contain the following additional information:

- N_A: the new AP's identity token, and
- ANonce: the random nonce generated by the new AP for computing a PTK.

Address Length	Status	MAC Address	Sequence Number
Octets: 1	1	n = Address Length	2

Figure 53: Cache-Response data field format

The exchange of FATP Requests and Responses between the Madwifi driver and the hostapd is supported by events (for driver to hostapd signalling) and by the ioctl interface (for hostapd to driver communications). On the client side, the wpa_supplicant application was modified to manage the FATP exchange and the FATP fast re-association procedure.

8.1.3 LM-SS Implementation

The LM-SS scheme is proposed in this research to reduce the number of channels to scan as well as eliminating channel waiting times associated with scanning. For the implementation of this scheme, the following three messages are introduced:

- **Scan Report**: contains a list of neighbour entries
- **Neighbour List Request**: used by a STA to request neighbour information from the location server
- **Neighbour List Response**: the reply to the above request to inform the STA about neighbour APs that are adjacent to the STA's current AP

These messages are transmitted as UDP/IP packets. Each message contains one header followed by zero or more neighbour entries. The message header and the neighbour entry use the format as shown in Figure 54 below.

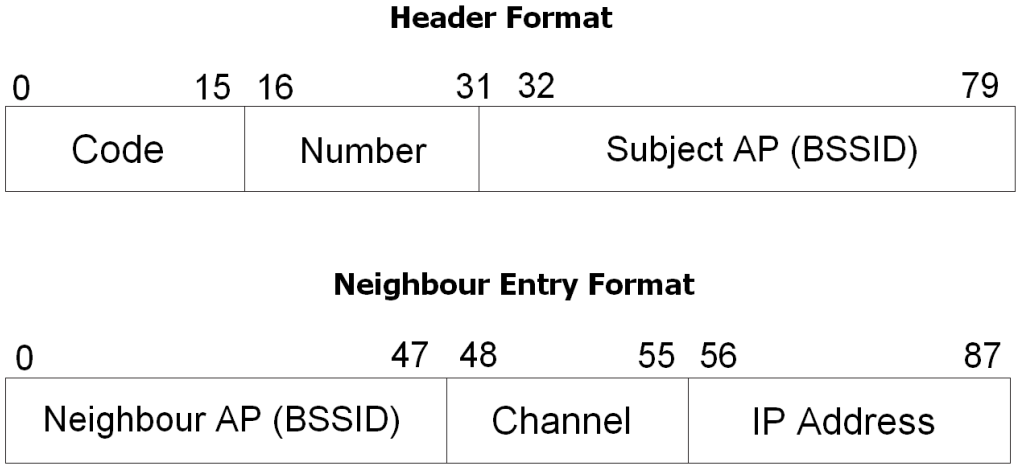

Figure 54: LM-SS message header and neighbour entry format

The two-octet Number field specifies the number of neighbour entries included in the message. The two-octet Code field is used to indicate the message type as the following:

- Code (0) – Scan report
- Code (1) – Neighbour List Request
- Code (2) – Neighbour List Response

With the LM-SS scheme, the original 802.11 probe request frame is extended to include an additional information element containing the IP address of the STA. The implementation of the LM-SS scheme also requires some modifications to the Madwifi driver, and a LM-SS application is also implemented on STAs to interact with the Madwifi driver as well as handling the exchange with the location server.

The Madwifi provides a utility tool called *iwevent*, which implements a user space 'IOCTL' to query the driver about different types of events, such as new associations with an AP and the completion of a scan process. In this implementation, the LM-SS application uses *iwevent* to capture new AP association events generated from the wireless driver. This is done by calling IOCTL with the command ID "SIOCGIWAP". With the acquired information from the driver, the LM-SS application is able to get the BSSID of the new AP and send the scan reports to the location server. Also with *iwevent*, whenever a scanning request is completed and results of the scan are available, an event is generated to signal the LM-SS application. Another utility in Madwifi called *athchans* is used to define the set of channels to scan for Madwifi devices. When a STA receives the neighbour list response from the location server, the LM-SS application will feed the information in the neighbour entries to the *athchans* to specify the channel numbers for the next scan. Both the client's and the server's LM-SS applications are implemented in C language and run in user space.

Fortunately the Madwifi version (0.9.4) used in the testbed supports the functionality of scanning in PS mode, which is referred to as "background scan" in Madwifi. This is modified so that when in PS mode, instead of passively listening for beacons from each channel, the STA actively sends out fast probing requests only to the neighbour APs.

After sending out fast probing requests to neighbour APs, the STA comes back to its original channel and waits to receive a beacon frame from its current AP. If the beacon frame indicates that the AP has frames queued, the STA exits PS mode, and pulls the buffered frames from the AP. When a fast probing response is received, the STA extracts the SNR value and stores it in the cached neighbour list. The neighbour AP with the highest SNR will be selected to perform FATP handoff when the current AP's signal strength degrades to a certain threshold.

8.1.4 Handoff Decision and Process

The 802.11 standard does not specify when a handoff should be triggered. Most of the existing implementations use the signal strength of the current AP as a decision making criteria. For the implementation of this research, the SNR of the AP is used to determine if a handoff is needed and when to start the scanning process, as shown in Figure 55.

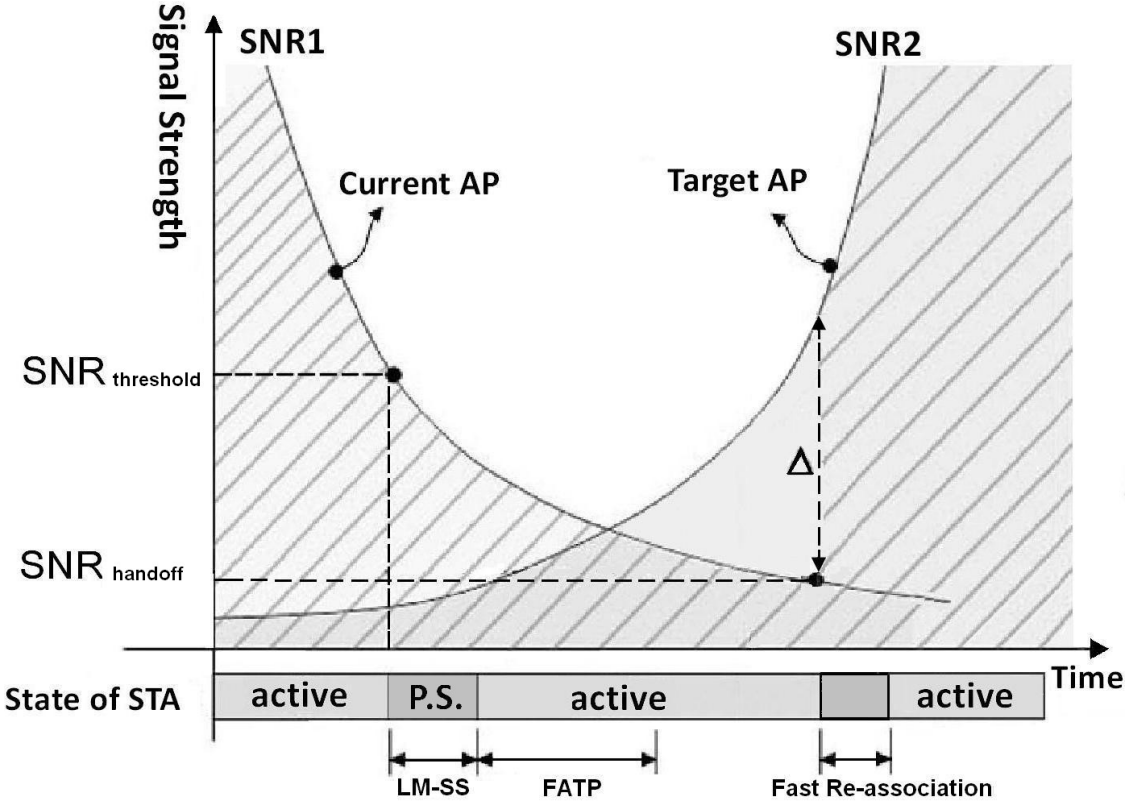

Figure 55: Handoff decision and hysteresis condition

In the figure, SNR1 and SNR2 indicate the signal strength of the current AP and a neighbour (target) AP, respectively. As the STA moves towards the target AP, the SNR of the current AP will decrease while the SNR of the target AP will increase. Periodically, the STA checks the SNR of its current AP. When the SNR is below a pre-defined threshold value, the decision to handoff is made.

Unfortunately, the signal strength of APs may change with respect to space, interference, and environment conditions. If only a single threshold value is used to determine the initiation of a handoff, it is likely to have a situation where the STA keeps initiating handoffs between APs because the SNRs of nearby APs are equally close to the threshold value. This effect is called the "ping-pong effect" [89]. To avoid this toggling, the testbed implementation uses a mechanism called hysteresis – handoff is allowed only when the signal strength of the new AP is better than the current one by at least a hysteresis constant, denoted as Δ. Thus, unnecessary handoffs can be reduced with the use of hysteresis. The hysteresis condition is also depicted in Figure 55 above.

As illustrated in the figure, when SNR1 drops below the $SNR_{threshold}$ value, a handoff is needed and the STA will start preparing for a handoff by initiating LM-SS and FATP. The actual handoff (i.e., disconnecting from the current AP and performing fast re-association with the new AP) will happen only when SNR1 further drops to a value, denoted as $SNR_{handoff}$ in the figure, which satisfies the following conditions:

$$\mathbf{SNR_{handoff} < SNR_{threshold}}, \text{ and}$$

$$\mathbf{SNR2 > SNR_{handoff} + \Delta}$$

The $SNR_{threshold}$ and the hysteresis constant are configurable in the implementation, and are currently set to 20db (approximately -80dBm RSS) and 7dB, respectively. These numbers were determined experimentally to give the best results.

8.2 Performance Evaluation

This section presents experiments and measurements performed in different scenarios to analyse, in detail, the performance and effectiveness of the proposed solutions in different situations.

8.2.1 APN Authentication Overhead

Instead of using Open System authentication, APN authentication is used during the initial stage of a RSNA establishment. A complete APN authentication involves a four-message exchange, as illustrated previously in Figure 22 on page 65. The authentication process comprises four main tasks that will incur delays: STA preparation (i.e., generating P, Q and N), puzzle construction at the AP, puzzle solving at the STA, and puzzle verification at the AP. Depending on the length of the identity tokens used in the exchange, the overall authentication latency can vary significantly. Generating primes requires computation time in the order of seconds, and hence, as a proof of concept implementation pseudo-primes are used to allow faster computation. The different length of identity token (N) and the corresponding latency of a complete APN authentication is summarised in Table 11.

Size of N (bits)	**Average computation time (ms)**				
	P	Q	N=P*Q	N|P	**Complete APN Authentication**
128	2.2857	2.2814	0.00117	0.00644	6.7355
256	**6.2895**	**6.2857**	**0.00158**	**0.00825**	**15.8480**
512	19.5731	19.5816	0.00194	0.01957	46.7564
1024	117.275	117.308	0.00293	0.02152	240.0821

Table 11: Identity token length and associated APN authentication latency

For the testbed implementation, 256-bit identity tokens are used because they provide enough key space to prevent cracking of validating keys within the required timeframe while the overall authentication latency is relatively short. Table 12 shows the delay components of a 256-bit N APN authentication. The results are the average of 20 authentications.

Complete APN Authentication Overhead (Based on 256-bit Identity Token)	
Task	**Time (ms)**
STA: Preparation	12.5736
AP: Puzzle Construction	0.0452
STA: Puzzle Solving	0.0129
AP: Puzzle Verification	0.0247
Transmission and RTT	3.1916
Total:	**15.848**

Table 12: Break down of 256-bit N authentication latency

From the above tables, it is obvious to see that the major source of delay comes from the computation of validating keys in the STA preparation phase, while the second comes from the frame transmissions and the Round Trip Time (RTT). Because of the lightweight-nature of APN authentication, the puzzle construction, solving, and verification can be done very quickly with trivial delays. With the 256-bit identity token based APN authentication, the average overall latency is about 15.85ms. Comparing this delay to the existing initial network connection time (i.e., performing full scanning plus 802.11/EAP-TLS) of almost 1.5 seconds, the APN authentication introduced only about 1% increase to the network connection time.

To evaluate the impact of APN authentication and frame validation functionality on the maximum AP bandwidth, the traffic generator in the testbed runs Iperf to generate different traffic loads from 2Mbps to 54Mbps to a receiving STA through an AP with

and without APN authentication and frame validation enabled. The result, as depicted in Figure 56, shows that the maximum bandwidth of the APs in the testbed can go almost up to 18Mbps; with the APN authentication scheme enabled, there is no noticeable performance degradation observed to the average maximum bandwidth.

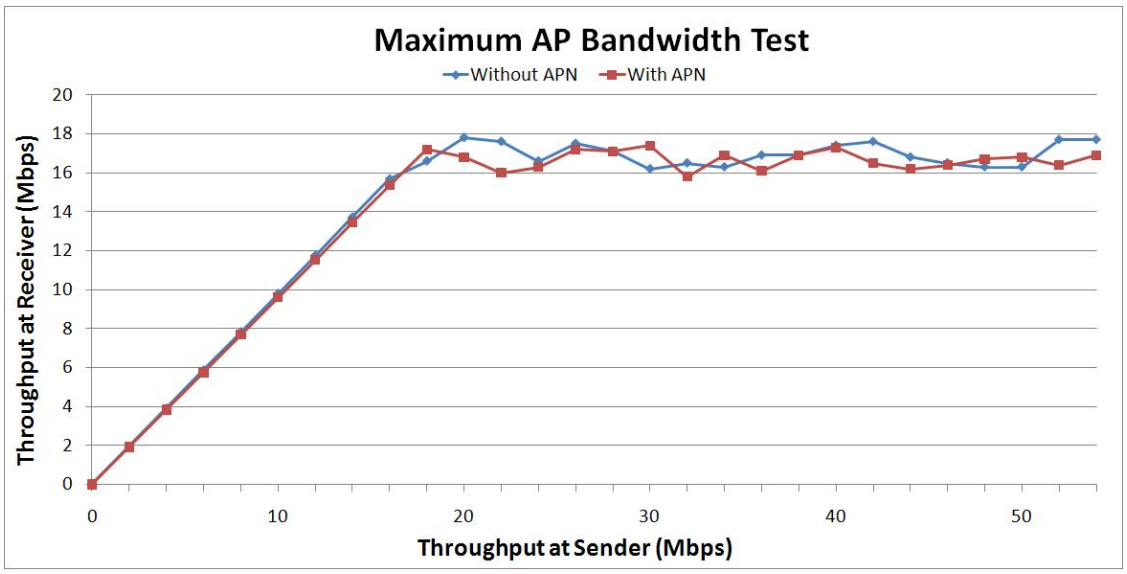

Figure 56: Maximum AP bandwidth with/without APN authentication scheme

8.2.2 DoS Mitigation Evaluation

To further quantitatively evaluate the effectiveness of the APN scheme in protecting against different DoS attacks targeting the STA and the AP at different stages of a RSNA establishment, various attack tools mentioned in Chapter 3 are used to launch the following attacks:

Prior to the initial 802.11 authentication and association:

- Authentication frame flooding
- Association frame flooding

After the secure communications link fully established:

- Deauthentication frame flooding
- Disassociation frame flooding

During 802.11i authentication and key establishment:

- EAPOL-Start flooding
- EAPOL-Logoff flooding
- EAP-Failure flooding

The experimental results are summarised in Table 13 below. The results demonstrated that the APN authentication scheme is effective in protecting against spoofing based DoS attacks.

DoS Attack Type	Without APN Scheme	With APN Scheme
Authentication frame flooding	AP resource depletion	Successful mitigation
Association frame flooding	AP resource depletion	Successful mitigation
Deauthentication frame flooding	STA connectivity loss	Successful mitigation
Disassociation frame flooding	STA connectivity loss	Successful mitigation
EAPOL-Start flooding	AP resource depletion	Successful mitigation
EAPOL-Logoff flooding	STA connectivity loss	Successful mitigation
EAP-Failure flooding	STA connectivity loss	Successful mitigation

Table 13: DoS attack results with/without APN authentication scheme

To examine the AP resource utilisation in handling the spoofed frames under a high rate flooding attack, an authentication flooding is performed at a rate at 18Mbps, which pushes the AP to the maximum bandwidth capacity, and the CPU and memory utilisation are monitored over the flooding period of 10 minutes.

Figure 57 shows the AP's CPU utilisation under the flooding condition. Without the APN authentication scheme, the CPU load is almost 80% under the flooding condition. This is because the AP was responding to the spoofed requests and allocating resources. On the other hand, with the APN authentication scheme enabled, the CPU utilisation is less than 40% under the same flooding condition. The significant reduction comes from the ability provided by the APN scheme to effectively identify spoofed frames

and discard them without storing any state information. Similar level of CPU load reduction was also found with other types of flooding attacks.

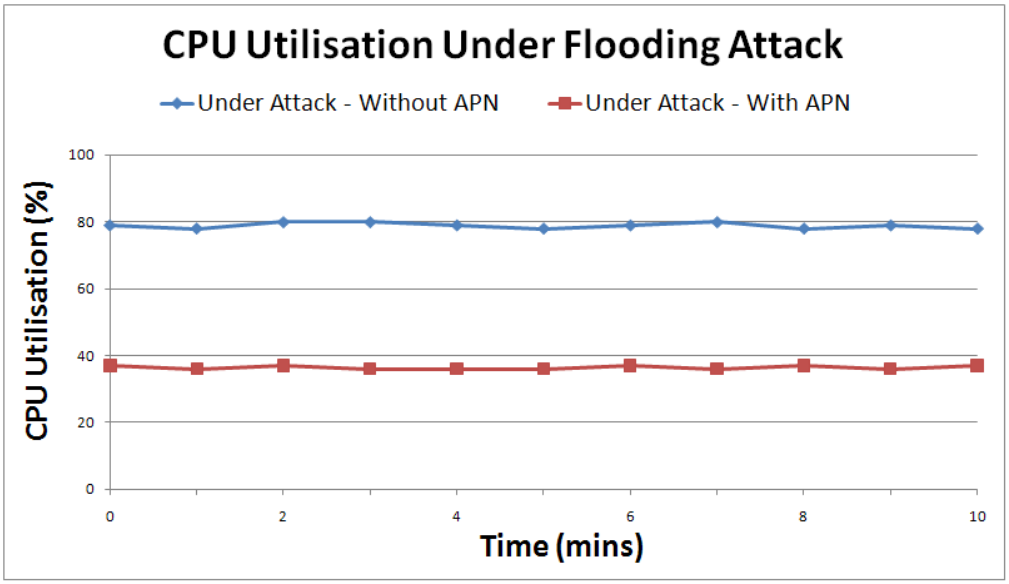

Figure 57: AP's CPU utilisation under flooding condition

Without the APN authentication scheme, the AP was not able to identify spoofed requests and actually responded to them, so the memory usage under the authentication flooding attack was continuing to increase, as depicted in Figure 58. In contrast, with the APN authentication scheme enabled, the AP's memory usage remained steady.

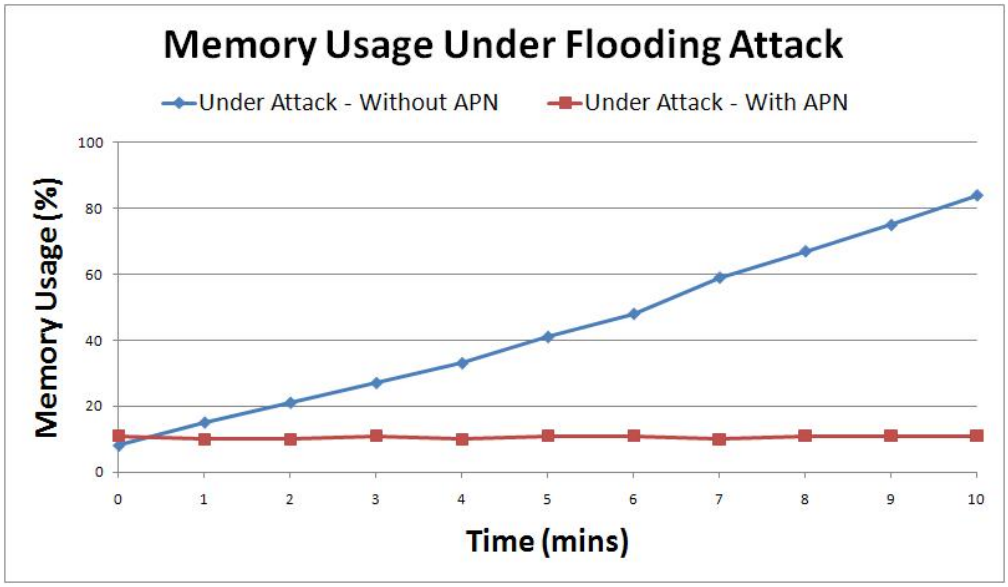

Figure 58: AP's memory utilisation under flooding condition

To evaluate the effect of DoS protection on the client side, the same deauthentication flooding attack previously performed in Section 4.2.2 was launched again with the APN authentication scheme enabled. The duration of attacks was 15 seconds. The following figure shows that without the APN authentication protection the STA was disconnected immediately right after the attack started, and the throughput remained zero over the duration of the attack. With the APN authentication scheme enabled, the STA throughput was not affected much because the spoofed frames were dropped without affecting other legitimate traffic.

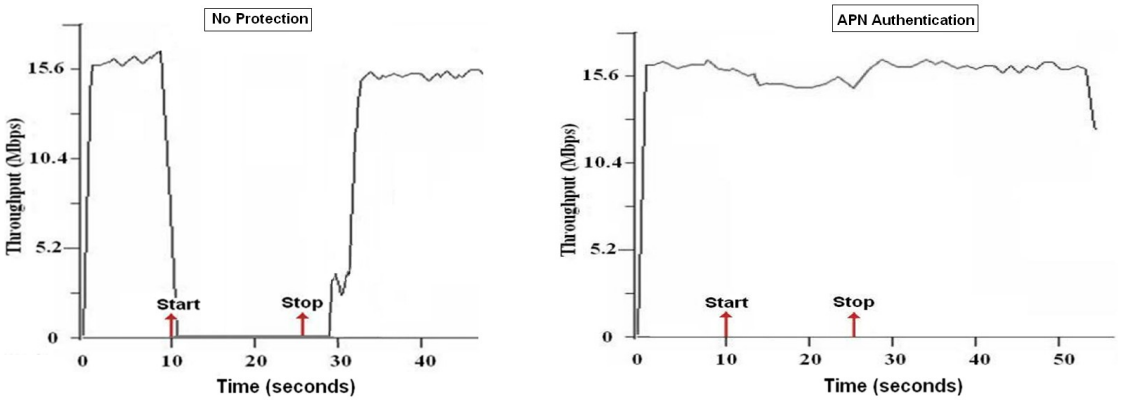

Figure 59: Throughput under deauthentication flooding with/without APN authentication protection

Another experiment to test the effect of the spoofed frame protection on the client side and its performance impact on real applications were also performed under flooding conditions. Both TCP and UDP application traffic were tested in the experiment. For this test, a large file was transferred from the traffic generator host to the STA while the AP was under attack during the file transfer. Both FTP (for testing TCP traffic) and TFTP (for testing UDP traffic) protocols were used to transfer the file. An EAP-Logoff flooding attack was launched 40 seconds after the start of the file transfer for a duration of 30 seconds. The results are plotted in Figure 60 and Figure 61.

In Figure 60, the TCP throughput at the receiving STA immediately dropped to zero after the attack was launched. This is because the spoofed EAP-Logoff frames were continuously disconnecting the victim STA and causing new associations to fail. The result further shows that not only the TCP session was disconnected during the attack,

but also the throughput continued to remain zero even after the attacked was stopped at the 70th second. This demonstrated that the consistent packet loss caused by the attack could actually severely disrupt TCP sessions and hence TCP based applications.

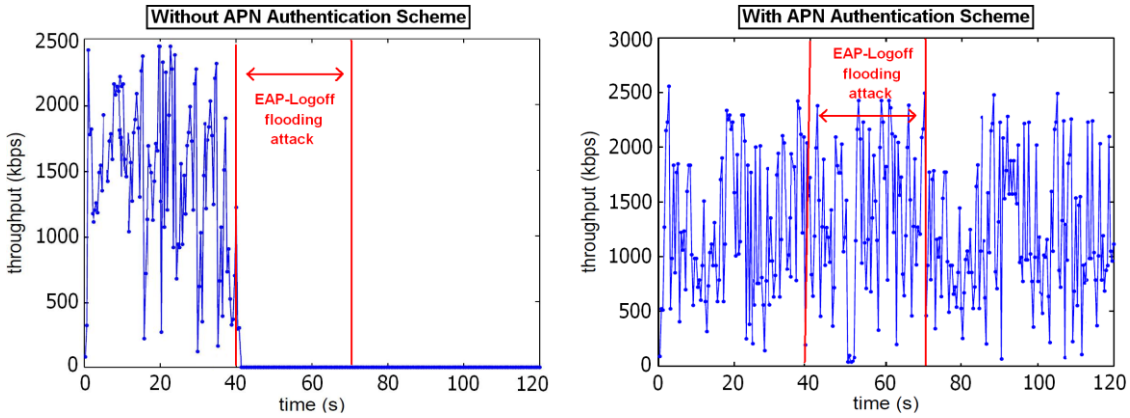

Figure 60: TCP throughput at the receiving STA under flooding

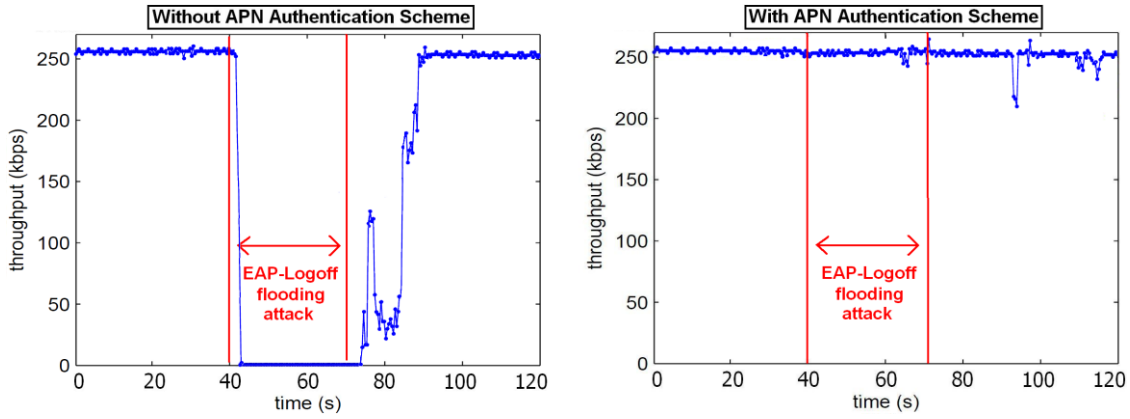

Figure 61: UDP throughput at the receiving STA under flooding

Figure 61 shows a similar result with the UDP traffic. However, the UDP traffic was able to resume its normal throughput after the attack was stopped because of the connectionless nature of UDP as well as the timeout and re-transmission mechanism of the TFTP protocol.

On the other hand, with the APN authentication scheme enabled, the frame verification was able to effectively filter out the spoofed EAP requests, thus the throughput at the receiving STA was not affected even though the AP was under a flooding condition.

With the experimental results obtained in this section, it is evident that the APN authentication scheme was able to effectively mitigate most of the spoofing based DoS and flooding attacks with minimal increase in the computation cost and traffic overhead. The lightweight and stateless nature of the scheme allows the AP to verify and discard spoofed frames, even under high flooding rate attacks, without becoming a new form of DoS vulnerability.

8.2.3 Handoff Performance

This section presents the experimental results of the proposed handoff solution, which combines the location management-based selective scan (LM-SS) and the FATP schemes. In each experiment, the client STA was moved from one AP to another to observe the handoff latencies and packet loss. The experiments were carried out at normal walking speed.

8.2.3.1 Scanning Delay

Scanning delay depends on the number of channels to scan and the time spent in each channel. In addition, there is always a constant overhead of Channel Switching (CS) delay, which is hardware implementation dependent. The channel switching delay is the time required for the WNIC to switch to a new frequency, resynchronise and start demodulating packets. Measurements were performed in the testbed network and it was found that the Atheros 5002 chipset based cards incur an active scanning delay of roughly 420ms (tested with Windows XP driver). This finding is consistent with the results reported by [71] and [84]. The channel switching delay of those Atheros cards was measured to be 5.25ms.

The scanning delay of the proposed LM-SS scheme can be determined by the following equation:

$$Scanning\ delay = CS \times (n + 1) + t_{probe} \times n$$

where *CS* is the channel switching delay, *n* is the number of channels to scan, and t_{probe} is the time required to send a fast probing request on a channel.

From the experimental measurements, the average t_{probe} was found to be close to 1.3ms. Hence, the expected scanning delay for scanning a certain number of channels can be calculated with the above equation, and the scanning delays are plotted in Figure 62.

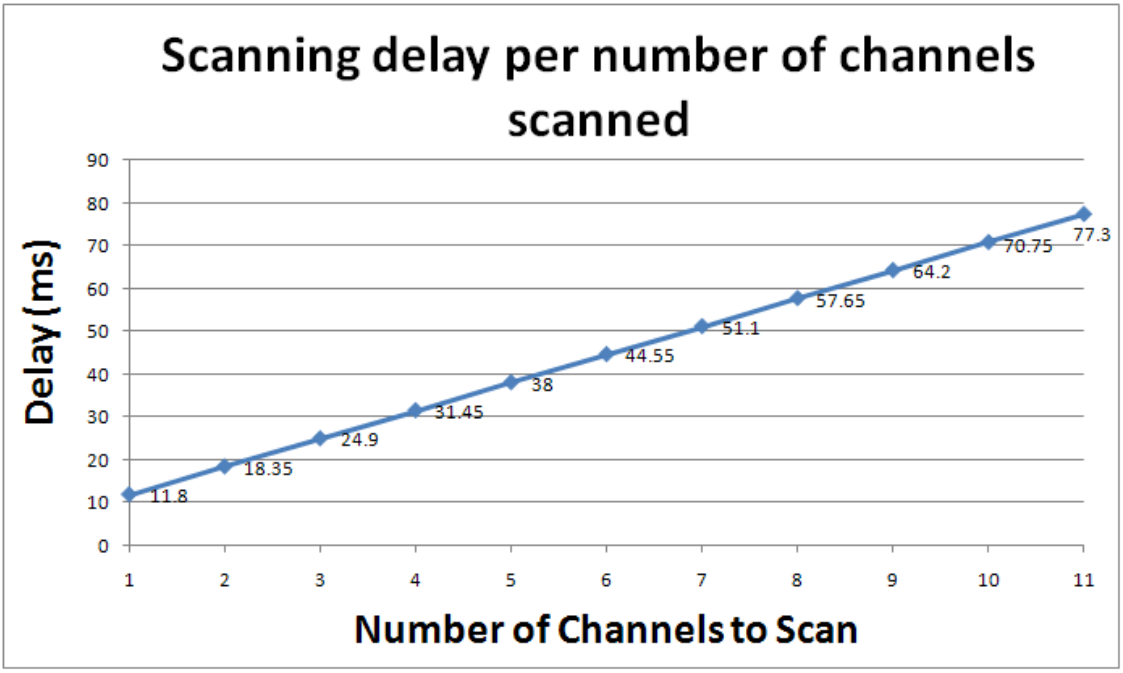

Figure 62: LM-SS scanning delay per number of channels scanned

The scanning delay is linearly depending on the number of channels to scan, so there will not be time wasted on probing channels that are empty. The specific channels that are required to probe are reliably provided by the location server after the STA associates to an AP.

The scanning delay also represents the communications outage time where the STA will not be able to receive the ongoing data traffic. In a typical enterprise WLAN where there are usually two or three neighbouring APs, the disconnected time due to scanning would be around 25 ms. In an extremely rare case where all 11 channels are used, the required scanning delay would be close to 80ms, which is still a huge reduction compared to the 420ms delay with the original active scanning.

Experiments were also conducted to compare the scanning delay under different AP density compared with different scan methods. Three scan methods were tested: 802.11 active scan, selective scan, and LM-SS. The selective scan is the time required to perform an active scan to probe only the channels used by neighbouring APs. The inclusion of the selective scan here has the benefit of using IP-based probe responses being able to be evaluated, as the LM-SS scheme without utilising IP based probe responses is equivalent to the selective scan. The following are the average of 10 runs of scanning and are summarised in Table 14.

Number of neighbouring APs	Scan Method	Average Delay (ms)
1 (channel 1)	802.11 Active Scan	417.82
	Selective Scan	55.46
	LM-SS	11.80
2 (channel 1 and 6)	802.11 Active Scan	417.82
	Selective Scan	112.75
	LM-SS	18.35
3 (channel 1, 6, and 11)	802.11 Active Scan	417.82
	Selective Scan	166.58
	LM-SS	24.90

Table 14: Average scanning delay with different scan methods

The results show that the active scan has the longest and fixed average delay. This is because all available channels were scanned without regard to the presence of neighbouring APs. Selective scan has a reduced average delay because the number of channels to scan is reduced to the number of neighbouring APs that are present. LM-SS further reduced the delay by using IP based probe responses as the waiting time in each channel is eliminated. Figure 63 graphically compares the results obtained under a network environment containing 3 neighbouring APs.

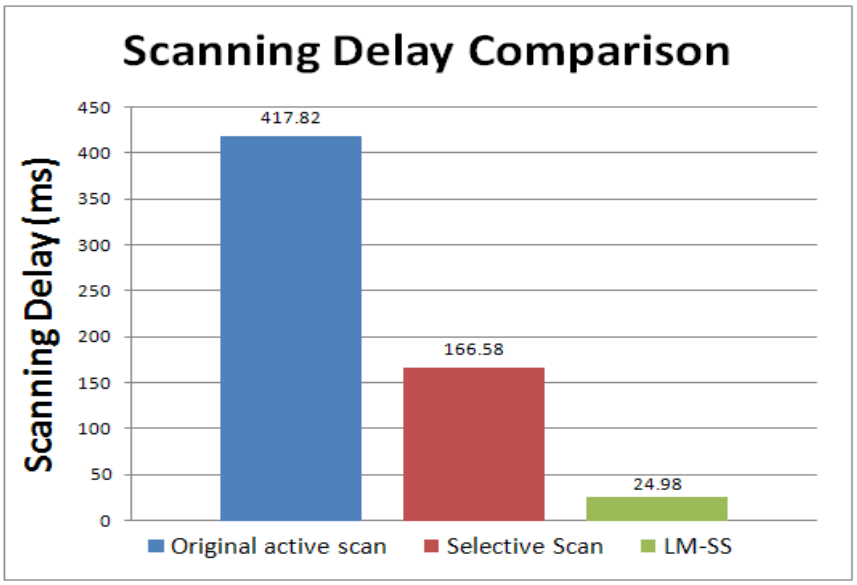

Figure 63: Three-channel scanning delay comparison

8.2.3.2 Handoff Latency

For the analysis of handoff performance, the term *(actual) handoff latency* is defined as the time period during handoff in which a STA cannot send and receive normal data packets due to the link-layer activities. To correctly determine the actual handoff latency, the tool *80211debug*, which is provided in Madwifi, is used to capture all important link-layer events from the driver and to provide additional, useful debug output for analysis.

Another term, *raw handoff latency*, is defined as the difference in the time at which the decision for a handoff is made, to the time at which the STA is successfully re-associated to the new AP. The raw handoff latency is as labeled in Figure 64, which depicts the handoff activities and the corresponding delays and the STA status.

The distinction between the two different measurements of latencies has been made because, with the proposed schemes, the raw handoff latency does not reflect the handoff performance, whereas the actual handoff latency actually represents the disconnection time from the data communications which directly affects the packet loss, and thus, QoS of the service.

With the proposed handoff solutions (LM-SS and FATP), the raw handoff latency ranges from 100ms to 180ms, depending on the number of channels required to scan. However, because of the proactive nature of the FATP scheme, the actual handoff latency only comprises the scanning delay and the fast re-association delay. Based on 10 handoff experiments performed in the testbed, the average fast re-association delay was measured to be 18ms. Hence, the possible range of the actual handoff latency would be between 30ms to 95ms, depending on the number of channels to scan.

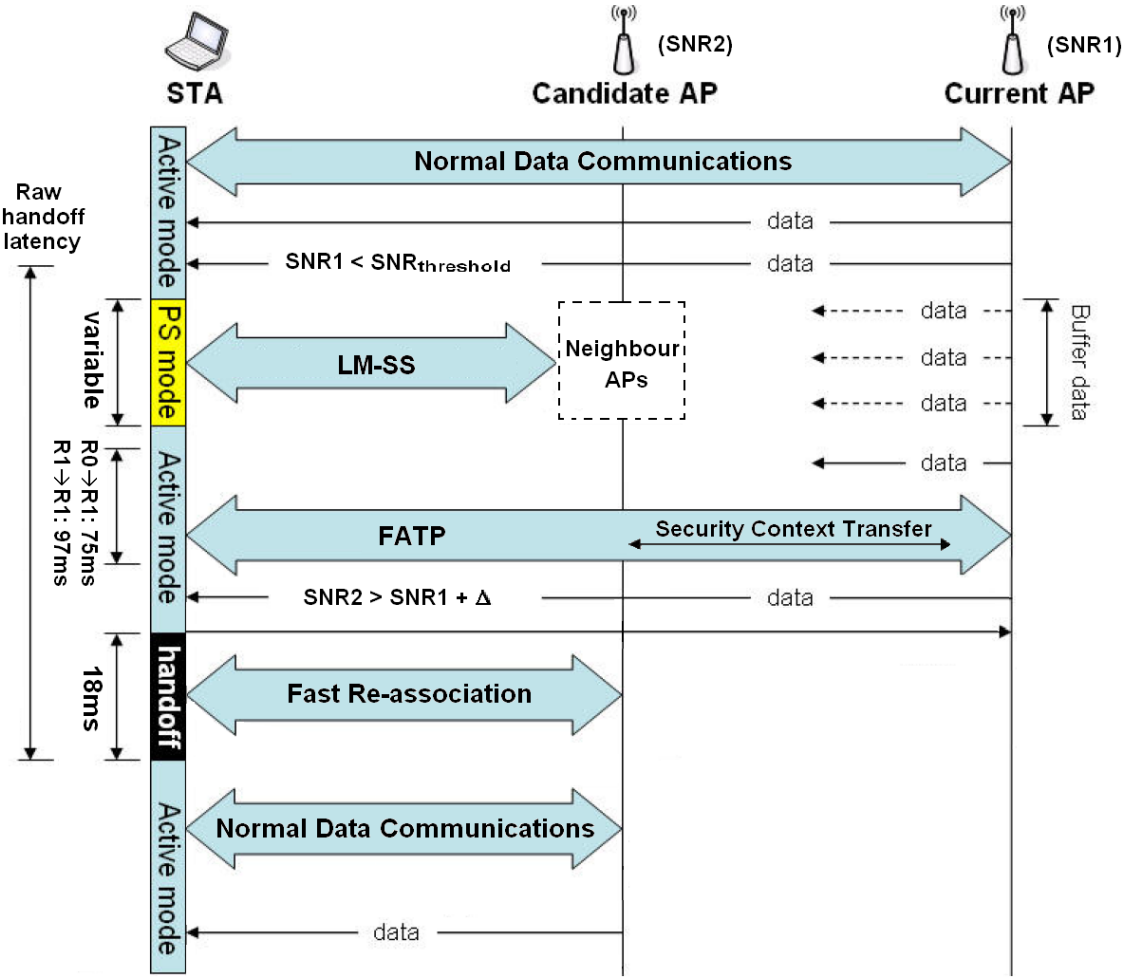

Figure 64: Handoff activities and the associated delays

The FATP trust establishment process on average takes about 75ms to complete for a R0→R1 handoff and 97ms for a R1→R1 handoff. The R0→R1 FATP delay includes a 28% (21ms) of IAPP communications overhead with the RADIUS server, whereas a R1→R1 handoff has roughly double the amount. The IAPP overhead involves activities such as AP identity verification, IP address lookup, security block exchange,

and security key distribution with the Cache-Notify and Cache-Response messages. The FATP delay and the IAPP overhead do not increase the actual handoff latency because the trust transfer and key distribution are performed prior to handoff while the STA is still connected to the current AP. Therefore, the actual handoff latency is independent of the type of roaming scenarios.

A typical enterprise WLAN environment usually utilises the 3 over-lapping channel cellular AP structure (refer to Figure 40). To evaluate the performance of the proposed handoff solutions under this environment, 10 handoffs were performed in a three neighbouring AP scenario in which the STA roams between four APs and is required to scan three channels for each handoff. The results are graphed in Figure 65.

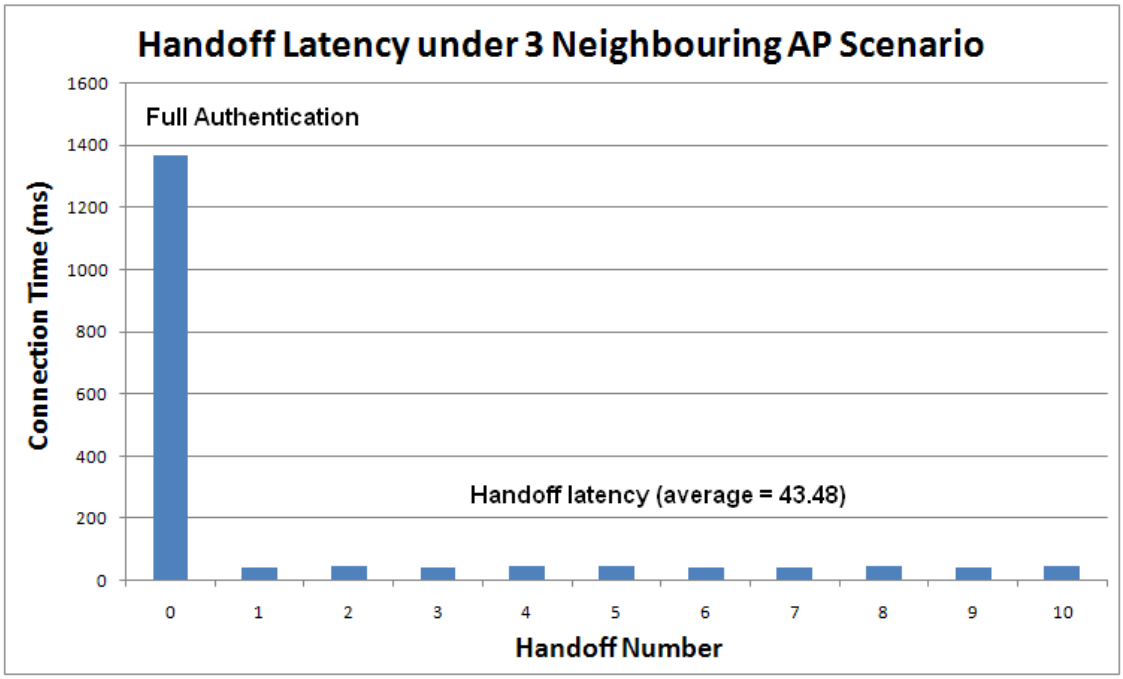

Figure 65: Handoff latency under three neighbouring AP structure

The initial network connection takes about 1.4secs to establish. This involves delays from the full active scan (420ms), re-association (25ms), full 802.11i authentication with EAP-TLS (850ms), and the key managements with four-way handshake and group key handshake (65ms). After that each handoff incurs an average latency of 43.48ms. This latency is small enough to achieve seamless mobility as well as ensuring the QoS of real-time multimedia services.

The performance of the FATP handoff scheme is compared with the existing 802.11i pre-authentication (using EAP-TLS). The following diagram shows the different components that constitute the actual handoff latency. The major portion of the delays comes from the scanning phase. With the LM-SS scheme, the result indicates that the scanning delay is significant reduced. This is because with FATP the pre-authentication and context transfer are done prior to handoff while the STA is still connected to its current AP. The FATP authentication delay does not contribute to the actual handoff latency, whereas the 802.11i pre-authentication incurs an authentication delay of around 350ms. The FATP scheme also has a reduced re-association delay because the four-way handshake is not required.

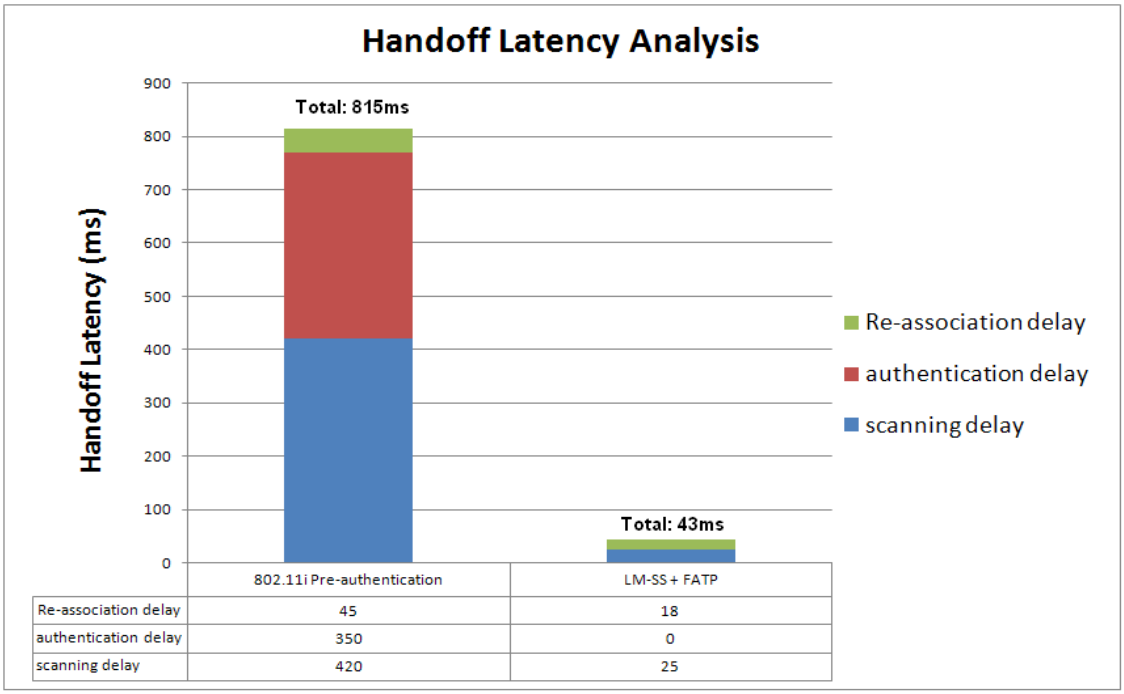

Figure 66: Handoff latency components and comparison

For this comparison, the scanning delay is excluded and only the (re)authentication delay of those handoff solutions are considered. The result is shown in Figure 67. With the original 802.11 handoff, a complete 802.1X/EAP-TLS authentication was performed, which resulted in the longest authentication delay. With the 802.11i pre-authentication, the delay was reduced to 350ms. This is because most of the authentication work was done through the current AP prior to the disconnection of the STA. With the 802.11r fast transition scheme, performance evaluations [90] have

shown that a fast transition requires 40 to 50ms (excluding scanning delays). The proposed FATP scheme only took 18ms, which is the time required to execute the fast re-association with the new AP.

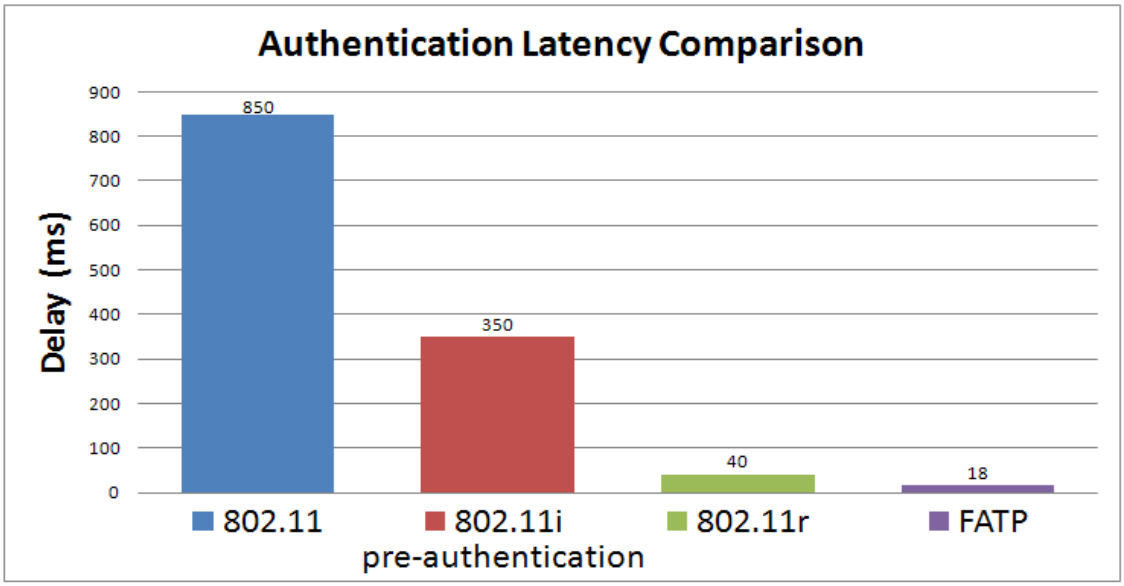

Figure 67: Handoff latency comparison between different handoff schemes

With the FATP, a roaming STA's context information has already been transferred to the new AP prior to disconnection, so the STA can directly re-associate with the new AP without needing to negotiate with the authentication server (RADIUS). This significantly reduces the authentication delay. Compared with the 802.11r, the further improved performance comes from the fact that with the FATP scheme the PTK security association between the new AP and STA has already been established locally prior to handoff, so the fast re-association phase does not need to perform the four-way handshake, whereas the 802.11r still requires a four-way handshake in order to complete the key setup.

8.2.3.3 Packet Loss during Handoff

The packet loss refers to the number of IP packets that are lost during a handoff. Packet loss due to a handoff is caused by the connection down time of the STA as well as the link-layer activities, such as channel scanning or re-authentication, which are required to perform a handoff. The delay of link-layer update in DS also incurs packet loss

because, by which time, the traffic is still routed to the old AP while the link to the old AP is already broken. The packet loss is calculated as the following:

Packet Loss = number of packets sent from the source – number of packets received at the destination

The number of packets lost directly reflects the service quality seen by the real-time applications, such as VoIP. In order to measure the packet loss during handoff, an UDP packet stream was transmitted from the traffic generator host in the testbed to a roaming STA. To simulate VoIP traffic, the source was sending a 200-byte packet[19] at an inter-arrival time (IAT) of 20ms. This is equivalent to an 80kbps data rate. Fifteen handoff experiments were performed in the testbed. For each experiment, the number of UDP packets transmitted by the generator and the number of packets received by the STA during the handoff period are recorded. The handoff period is the time between when the STA sends the first probe request and when the STA receives the first UDP packet through the new AP. The numbers of the packets sent and received were recorded accumulatively. The results are shown in Table 15.

With the original 802.11 standard, a handoff completely obstructed the ongoing communications during the handoff period, and thus resulted in a 100% packet loss rate. With the 802.11i pre-authentication, the STA disconnection time was reduced, so the packet loss rate decreased to 39.9% as a result. Almost 90% of the packets lost here were caused by the conventional active scan, which always incurs a long delay.

With the LM-SS + FATP solution, the packet loss rate was found to be less than 1%. This significant improvement in packet loss is the direct result of packet buffering

[19] Typically the data payload size of VoIP is 160 bytes in addition to the 20 byte IP header, the 12 byte RTP header and the 8 byte UDP header.

provided by the power-saving mechanism of legacy 802.11, which successfully prevented packet loss while the STA was scanning for APs on other channels. The other contributor to the low packet loss is the short handoff latency with the FATP scheme. Apart from the scanning phase, the only obstruction to the STA's data communications is the fast re-association phase. As the delay associated with the fast re-association phase (18ms) is actually shorter than the packet IAT of the VoIP traffic (20ms), no obvious packet loss was observed as a result.

Handoff Scheme	802.11		802.11i Pre-authentication		LM-SS + FATP	
	#Packets sent	#Packets received	#Packets sent	#Packets received	#Packets sent	#Packets received
Run1	67	0	57	38	6	6
Run2	65	0	53	34	8	8
Run3	69	0	54	29	7	7
Run4	68	0	62	38	9	8
Run5	70	0	51	36	8	8
Run6	68	0	54	30	9	9
Run7	68	0	56	38	7	7
Run8	66	0	51	27	7	7
Run9	72	0	58	29	8	8
Run10	70	0	57	33	8	8
Run11	71	0	56	31	6	6
Run12	67	0	54	34	9	9
Run13	70	0	51	29	7	7
Run14	71	0	52	31	8	8
Run15	67	0	56	37	9	9
Total	1029	0	822	494	116	115
Packet Loss	1029 (100%)		328 (39.90%)		1 (0.86%)	

Table 15: Handoff packet loss of 15 successive handoffs

Therefore, by combining the proposed LM-SS scheme and the FATP scheme, the product is a secure and seamless WLAN handoff solution that is able to meet the QoS requirements of real-time applications.

8.2.3.4 VoIP Quality Evaluation

To put the proposed handoff solution into test for the actual handoff impact on VoIP services, a VoIP call is setup between a roaming STA and a wired client in the DS, and experiments were performed to observe the packet IAT at the STA. For this experiment, an Atheros 5002 WNIC commercial driver for Windows XP was compared with the LM-SS + FATP handoff solution implemented on Linux. The experiment was performed in the testbed running three APs operating at channel 1, 6, and 11.

Figure 68 below shows the comparison of the measured VoIP IAT affected by handoffs. The performance of the existing handoff mechanism with the commercial driver is depicted in the plot on the left hand side. A slight degradation in call quality due to poor signal strength was observed a couple of seconds before the handoff occurred at the 19th second. During the handoff, a long service disconnection time of around 5 seconds was noticed, and there were no voice packets received during this period, as indicated by the large IAT shown in the plot.

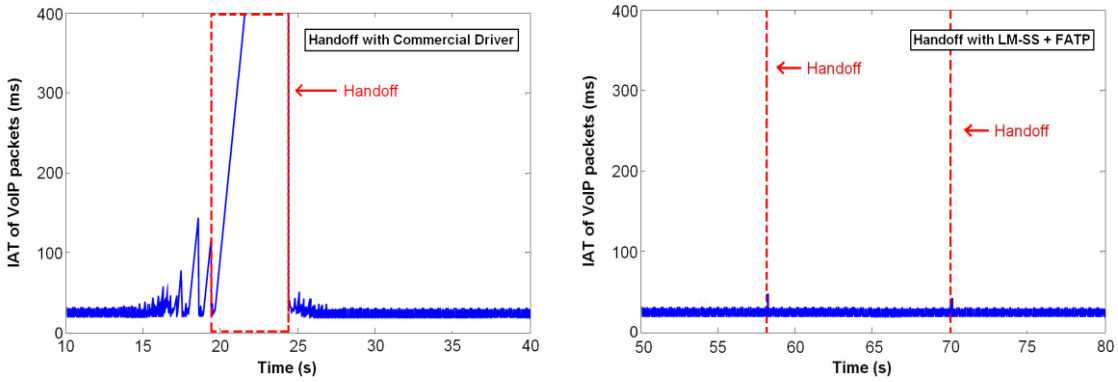

Figure 68: Handoff impact on VoIP IAT

The handoff performance with the proposed LM-SS + FATP handoff solution is shown in the plot on the right. Two handoffs were performed at the 58th second and the 70th second. There was no noticeable impact on the voice traffic before handoffs occurred because the scanning and the context transfer were done while the STA was still having a reasonably good link quality with its current AP (see Figure 69). During the first handoff, only a 45ms IAT for one packet was introduced. This was due to the

scanning phase, as the packet was buffered at the AP while the STA was sending fast probing requests on other channels, and delivered later to the STA. An IAT of 41ms for one packet in the second handoff was also noticed due to the same cause.

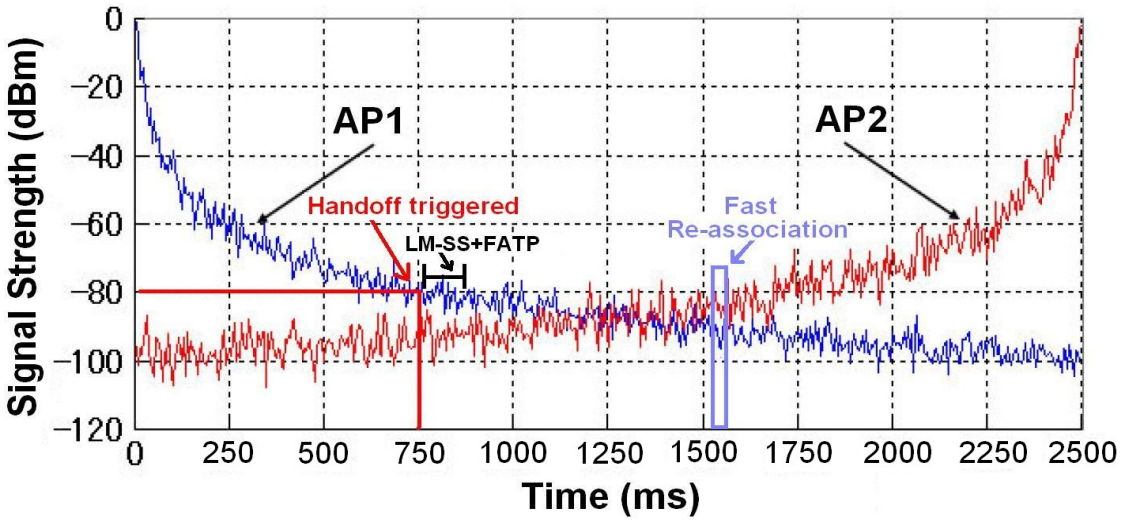

Figure 69: Handoff timing and the corresponding signal strength

Similar experiments were performed to also evaluate the effect of an increased number of neighbouring APs, and there was no noticeable impact to the voice quality found. The reason is as explained in the following:

The total service interruption time with the proposed handoff solution is the scanning delay (with the possible values shown in Figure 62) plus the fast re-association delay of 18ms. Because of the packet buffering mechanism, LM-SS does not incur any packet loss during the scanning phase. Thus, any packet that happens to arrive during the fast re-association delay will result in packet loss. Since the typical IAT of VoIP packets is 20ms, the FATP scheme will have at most one packet loss. The only factor that can degrade the handoff performance is the scanning delay, which is increased as the number of neighbouring APs increase. For example, a handoff involving scanning five neighbouring APs introduced an IAT of 57ms for one packet plus at most one packet loss during the fast re-association phase. From the experiment performed, the voice quality degradation because of this was not detectable by hearing. Further increase in the scanning delay will definitely incur noticeable impact on the traffic. However, it would be very rare for a practical enterprise WLAN to have such a high AP density.

8.3 Chapter Summary

The results presented in this section demonstrated the capability of the handoff solution using the proposed LM-SS and the FATP schemes in supporting fast and seamless link-layer handoffs in WLANs. Uninterrupted VoIP traffic during handoffs has been made possible with the significantly reduced handoff latency and low packet loss rate (<1%) provided by the proposed schemes.

Channel probing and scanning has always been the most time consuming activity in a handoff. With the key hierarchy of the 802.11i standard, the re-computation of new PMKs, PTKs, and key handshakes are also the major delay components constituting the handoff latency. The combined total delay is always unacceptable for multimedia applications. With the LM-SS + FATP handoff solution, most of these activities are carried out prior to the handoff and with minimal impacts on the ongoing data traffic. Thus, the service disconnection time of STAs is significantly reduced, resulting in a shorter latency and better VoIP quality.

Table 16 summarises the overall performance of related handoff schemes. Regarding the handoff latency, the proactive key distribution solutions, 802.11r roaming, and the proposed handoff solution can meet the latency requirement of multimedia applications. However, the proactive key distribution solutions require costly prediction algorithms as well as large traffic overhead in distribution keying information. Therefore, this type of handoff scheme is not scalable and thus not suitable for enterprise WLANs. The proposed handoff solution has a slightly better handoff performance compared to 802.11r. This improvement is the direct result of the simplified re-association procedure without the four-way handshake.

Combining all the desirable properties in terms of the handoff performance (latency, packet loss, and re-authentication delay), computation and traffic overhead, and security, the proposed LM-SS + FATP handoff solution appears to be the most attractive handoff implementation for single-domain enterprise WLAN environments.

Handoff Scheme Type	Handoff Latency	Packet Loss	Overhead	Data Buffering	DoS and spoofing prevention
Full authentication	Highest	High	Full RSNA overhead	No	No
Full authentication with lightweight EAP	Medium	High	Re-authentication overhead is unacceptable for handoff	No	No
802.11i Pre-authentication	Medium	High	High message overhead among STA, AP and AS	No	No
PKD	Small	Medium	Require costly prediction algorithms and large key pre-distribution message overhead	Yes (with some implementations)	No
802.11r	Small	Low	Affordable keying message overhead. Small storage space is required	No	No
LM-SS + FATP	Smallest	Lowest (<1%)	Considerate IAPP keying message overhead, but performed prior to handoff. Small storage space is required	Yes (using legacy 802.11 PS mode)	Yes (with APN authentication support)

Table 16: Performance comparison of handoff schemes

Chapter 9

Conclusion and Future Work

This chapter provides a summary of the overall research work and the proposed solutions presented in the thesis. The limitations of the proposed solutions are discussed, and the potential areas for future work are also suggested in this chapter.

9.1 Research Summary

IEEE 802.11i security standard provides an enhanced user authentication and strong data confidentiality to WLANs. However, the standard only concerns the protection of higher-layer data, i.e., IEEE 802.11 data frames, and the management frames used for connection administration are left unprotected. Hence, there is a wide spectrum of known attacks that are threatening to the WLAN security, particularly the DoS attacks. Mitigation solutions for wired networks have been widely researched and studied. As an example, one of the thesis author's previous works published in [48] proposed an effective DoS mitigation solution for SIP systems. However, link-layer DoS attacks in WLANs have not been fully mitigated to a satisfactory level. Although the IEEE 802.11w amendment was later introduced to further extend the data protection to the management frames, the experimental results in this research showed that the 802.11w protection is incapable of providing protection without causing severe performance degradation to the network under high rate flooding, and not all of the management frames can be protected with the 802.11w standard.

This research first studied the security of IEEE 802.11i amendment and identified some of the common link-layer DoS vulnerabilities. Experimental evaluations were performed to quantitatively measure the performance impact by the DoS attacks that exploit those vulnerabilities. Mitigation requirements were analysed and some

potential techniques to prevent spoofing and flooding activities were also discussed. Based on the results of this analysis, the thesis first proposed a lightweight, stateless frame authentication scheme, called APN authentication as introduced in Chapter 5 to address those DoS vulnerabilities. A RSNA can be established using the APN authentication instead of the existing Open System authentication, which does not actually provide any security. The APN authentication has the advantages of simplicity, compatibility with the standard with few modifications, and low computation overhead and bandwidth utilization. Experimental results demonstrated that with a small increase in the initial connection time, the APN authentication scheme is able to effectively identify between the legitimate traffic and the spoofed traffic, with a minimal computation cost due to the efficient client puzzle verification mechanism being used. This frame verification process allows the AP to quickly drop the spoofed frames under flooding conditions without affecting legitimate users' application traffic.

The research further focuses on improving the existing handoff performance in order to achieve secure and seamless link-layer handoffs that can meet the QoS requirements of real-time multimedia applications. The handoff performance is improved by shortening both the re-authentication latency and the channel scanning delays in the discovery phase. To achieve a secure roaming between APs with reduced re-authentication latency, a handoff scheme called Fast AP Transition Protocol (FATP) is proposed in Chapter 6. The FATP is a proactive key distribution based handoff scheme which delivers new session keys from the STA's current associated AP to the target AP prior to handoff. In order to achieve the same level of security as the 802.11i standard, the three-tier key hierarchy used in the IEEE 802.11r is adopted in the FATP scheme for managing and associating new security keys for handoff sessions. One of the major differences between the proposed FATP solution and the IEEE 802.11r roaming standard is the early execution of the four-way handshake prior to handoff with the FATP. This allows the FATP to achieve a faster re-association with the new AP, and thus, the handoff latency and packet loss can be further reduced compared to the performance of the 802.11r roaming. More importantly, the FATP supports the re-generation of the security parameters used by the APN authentication so that the subsequent link-layer frames can still be authenticated using the refreshed keying

material in new handoff sessions. The experimental results showed that the FATP scheme can achieve considerable reduction in handoff latency whilst providing the same security level as a full 802.1X authentication.

To achieve faster scanning than the existing 802.11 active scan, a location management based selective scanning (LM-SS) scheme was proposed in Chapter 7. In the LM-SS scheme, a location server is introduced to maintain the AP topology information, and a roaming STA can be informed with the presence of nearby APs, and therefore, determine the number of channels required to scan from the AP topology information provided by the location server. To eliminate the waiting time spent in each channel scanned, a cross-layer probing technique utilising unicast probe requests and IP-based probe responses is introduced. The packet buffering mechanism provided by the existing 802.11 power saving mode is used as a signalling mechanism for buffering packets while the STA is performing probing activities on other channels. This prevents packet loss during the scanning process. The experiment results, which are presented in Chapter 8, demonstrated that the combination of the LM-SS and FATP schemes is a complete suite of handoff solutions that can provide promising results to the VoIP user experience. Furthermore, with the frame protection provided by the APN authentication scheme, the RSN infrastructure can effectively mitigate most of the DoS attacks while having a greatly enhanced handoff performance.

9.2 Limitations

The design and implementation of the proposed handoff solutions in this research mainly concern intra-domain handoffs. Further research work, as suggested in Section 9.3, will be required to fully extend the FATP scheme to support handoffs between multiple domains. The APN authentication and the FATP handoff scheme both require the STA to generate prime numbers for the computation of identity tokens and validating keys. The computation time for generating large pure primes is normally in the order of seconds, which is too large for handoffs. Because of this, the implementation actually uses pseudo-primes instead. The optimisation of large prime computation will significantly improve the efficiency of those schemes.

Another weakness of the FATP scheme is that it does not fully conform to the 802.11i trust model, in which only the AS is trusted. The FATP scheme relies on the STA's initial AP (R0-AP), which authenticated the STA using the full 802.1X process, to generate and deliver new security keys for the handoff session. If a R0-AP is compromised, all the subsequent new PMKs are also compromised. To prevent this, additional enhancements to the AP's physical security and access control would be necessary.

The effect of the speed of mobility on the handoff performance has not been studied. The implementation of the proposed handoff scheme assumes a normal walking speed when handoffs are performed. If the STA is moving at a faster speed, the device may not have enough time to complete the LM-SS and the FATP trust transfer prior to the disconnection with the current AP, and therefore a full scanning and 802.1X authentication could result.

The major limitation of the LM-SS scheme is the inflexibility of the selection of the signal strength threshold value that triggers the scanning and the FATP. The SNR threshold and the hysteresis values specified in Section 8.1.4 may be only suitable in a small testbed environment. The practical threshold value and the hysteresis constant may vary in different enterprise WLAN deployments because they could be affected by the AP density, distance, signal interference, etc. If the threshold value is too small, it is possible that the STA will loss its connectivity with the current AP before the scanning and the security context transfer could be completed. Therefore, the optimised values need to be determined in the actual deployment.

9.3 Future Work

One of the major areas for potential future work is to extend the FATP to support inter-domain handoffs in WLANs. The proposed LM-SS and FATP schemes already provide some of the capabilities that will assist inter-domain handoffs. First, the neighbour reports from the location server contain the IP address used by the nearby APs. This will allow the STA to determine if an inter-domain handoff is required by

examining the IP address of the target AP. Second, the security context information is transferred over the DS using IAPP packets, and the same IAPP transfer mechanism can also be used for inter-domain handoffs, provided that a domain routing service, such as Mobile IP, is available in the DS infrastructure. However, a few issues still need to be considered and investigated in order to support inter-domain handoffs with the FATP:

1. Extending IAPP for inter-domain mobility requires enhancements to the security infrastructure with mechanisms for delivering authentication context information from the home AS to other domains;
2. The optimisation of communications between link-layer and network-layer is necessary in order to reduce delays associated with the cross-layer communications and overhead during handoffs;
3. A mechanism to effectively set up the roaming agreement among different domains needs to be provided. An investigation of how to securely share the new session keys issued from different domains is also needed;
4. To thoroughly evaluate these schemes quantitatively, a testbed implementation of the solutions is required to allow the evaluation to be performed in a well-defined, cross-layer model that can provide realistic performance indications.

Another area for future improvements is the new AP selection mechanism used in the LM-SS scheme. In some cases it is possible that the AP with the best signal strength indication may not necessarily be the best target AP for a handoff. This is because the channel used by the AP with strong signal strength is usually more congested under a higher density of STAs; the APs with moderate signal strength could actually be a better handoff candidate due to the lower traffic loading. Hence, it is important to also consider the loading of the APs when a STA is looking for a handoff candidate AP. As the location management scheme is extended to further provide the loading indication of the neighbouring APs to a roaming STA, the load balancing will certainly provide more values to the QoS of the end users' applications.

References

[1] "IEEE Std 802.11 - Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications," 1999.

[2] I. M. Scott Fluhrer, and Adi Shamir, "Weaknesses in the Key Scheduling Algorithm of RC4," *Eighth Annual Workshop on Selected Areas in Cryptography*, vol. Springer-Verlag, 2001.

[3] J. R. Walker, "Unsafe at any key size; An analysis of the WEP encapsulation," *IEEE 802.11-00/362*, 2000.

[4] "IEEE 802.11i-2004," *http://standards.ieee.org/getieee802/802.11.html*.

[5] C. Liu and J. T. Yu, "Review and Analysis of Wireless LAN Security Attacks and Solutions," *Journal of International Engineering Consortium*, vol. 59, 2006.

[6] S. Grech and J. Nikkanen, "A Security Analysis of Wi-Fi Protected Access," *The 9th Nordic Workshop on Secure IT-systems, Helsinki University of Technology, Finland*, 2005.

[7] W. Ge and S. Sampalli, "A Novel Scheme For Prevention of Management Frame Attacks on Wireless LANs," 2005.

[8] J. H. P. Ding, A. Celik, "Improving the Security of Wireless LANs by Managing 802.1X Disassociaion," *Proceedings of the IEEE Consumer Communications and Networking Conference, Las Vegas, NV*, pp. 53-58, 2004.

[9] C. Liu and J. T. Yu, "Protecting Enterprise Wireless LANs Using an Integrated Security Approach of VPN over 802.11i," *3rd International Conference on Cyber Information Technology and System Applications (CITSA), Orlando Florida*, pp. 278-283, 2006.

[10] J. I. A. Stubble eld, and A. Rubin, "Using the Fluhrer, Mantin, and Shamir Attack to Break WEP," *Technical Report TD4ZCPZZ, ATT Labs*, 2001.

[11] A. Stubblefield, Ioannidis, J., and Rubin, A., "Using the Fluhrer, Mantin, and Shamir attack to break WEP," *In Proceedings of the 2002 Network and Distributed Systems Security Symposium*, pp. 17-22, 2002.

[12] "Anton T. Rager. WEPCrack, http://wepcrack.sourceforge.net/."

[13] "Airsnort, http://airsnort.shmoo.com/."

[14] W. A. A. Jon Edney, "Real 802.11 Security: Wi-Fi Protected Access and 802.11i," pp. 107-108, 2003.

[15] Changhua He and J. C. Mitchell, "Security analysis and improvements for IEEE 802.11i," *Network and Distributed System Security Symposium Conference Proceedings*, 2005, http://www.isoc.org/isoc/conferences/ndss/05/proceedings/papers/NDSS05-1107.pdf.

[16] Pack S, Choi J, Kwon T, and C. Y, "Fast handoff support in IEEE 802.11 wireless networks," *IEEE Communications Surveys and Tutorials*, vol. 9, pp. 2-12, 2007.

[17] Baber Aslam, M Hasan Islam, and S. A. Khan, "Pseudo Randomized Sequence Number Based Solution to 802.11 Disassociation Denial of Service Attack," *in Proceedings of the First Mobile Computing and Wireless Communication International Conference*, pp. 215-220, 2006.

[18] L. B. a. J. Vollbrecht, "PPP EXtensible Authentication Protocol (EAP)," *IETF RFC 2284*, 1998.

[19] S. W. C. Rigney, "Remote Authentication Dial In User Services (RADIUS), RFC 2865," 2000.

[20] "RFC 3588 - Diameter Base Protocol. Network Working Group," 2003.

[21] T. W. W. Xu, W. Trappe, and Y. Zhang, "Channel Surfing and Spatial Retreats: Defenses Against Wireless Denial of Service," *In ACM WiSe*, pp. 80-89, 2004.

[22] W. Xu, W. Trappe, and Y. Zhang, "Channel Surfing: Defending Wireless Sensor Networks from Interference," *In IPSN*, pp. 499-508, 2007.

[23] S. S. J. Bellardo, "802.11 Denial-of-Service attacks: real vulnerabilities and practical solutions," *In Proceedings of the 12th USENIX Security Symposium, Washington, D.C.*, pp. 15-28, 2003.

[24] Aristides, Mpitziopoulos, and D. Gavalas, "An effective defensive node against jamming attacks in sensor networks," *Security and Communication Networks*, vol. 2, pp. 145-163, 2009.

[25] Mike Lynn and R. Baird, "Advanced 802.11 Attack. Black Hat Briefings," July 2002.

[26] R. Floeter, "Wireless Lan Security Framework: void11," *http://www.wlsec.net/void11/*, 2002.

[27] Y. Zhimin, C. C. Adam, G. Boxuan, B. Xiaole, and X. Dong, "Link-layer protection in 802.11i WLANS with dummy authentication," in *Proceedings of the second ACM conference on Wireless network security*. Zurich, Switzerland: ACM, 2009.

[28] Yixin Jiang, Chuang Lin, and Z. Chen, "A mutual authentication and privacy mechanism for WLAN security," *WIRELESS COMMUNICATIONS AND MOBILE COMPUTING*, vol. 8, pp. 101-112, 2008.

[29] F. A. Z. I. Martinovic, A. Bachorek, C. Jung, and J. B. Schmitt., "Phishing in the Wireless: Implementation and Analysis," *In Proceedings of the 22nd IFIP International Information Security Conference (SEC 2007). Springer LNCS*, 2007.

[30] S. E. Muthukkumarasamy, "Detecting Security Threats in Wireless LANs Using Timing and Behavioral Anomalies," 2005.

[31] H.-T. Chien, "Using Random Bit Authentication to Defend IEEE 802.11 DoS Attacks," in *Department of Computer Science and Information Engineering*: National Chiao-Tung University, 2006.

[32] E. Sithirasenan and V. Muthukkumarasamy, "Detecting Security Threats in Wireless LANs Using Timing and Behavioral Anomalies," presented at Networks, 2007. ICON 2007. 15th IEEE International Conference on, 2007.

[33] K. Byoung Uk, A.-N. Youssif, F. Samer, H. Salim, and Y. Mazin, "Anomaly-based fault detection in pervasive computing system," in *Proceedings of the 5th international conference on Pervasive services*. Sorrento, Italy: ACM, 2008.

[34] Chibiao Liu and J. Yu, "A Solution to WLAN Authentication and Association DoS Attacks," *IAENG International Journal of Computer Science*, vol. 34, 2008.

[35] B. M. Ferreri F, "Access points vulnerabilities to DoS attacks in 802.11 networks," *WIreless Communications and Networking Conference*, vol. 1, pp. 634-638, 2004.

[36] J. C. Changhua He, Mitchell, "Analysis of the 802.11i 4-way handshake," *In Proceedings of the 2004 ACM workshop on wireless security, ACM Press, New York, USA*, pp. 43-50, 2004.

[37] F. D. Rango, D. C. Lentini, and S. Marano, "Static and Dynamic 4-way Handshake Solutions to Avoid Denial of Service Attack in Wi-Fi Protected Access and IEEE 802.11i," *EURASIP Journal of Wireless Communication Networks*, vol. 2, 2007.

[38] T. D. D. M. Hannikainen, M. Niemi, and J. Saarinen, "Trends in Personal Wireless Data Communications," *Computer Communications*, vol. 25, pp. 84-99, 2002.

[39] C.-M. Huang and J.-W. Li, "A Context Transfer Mechanism for IEEE 802.11r in the Centralized Wireless LAN Architecture," presented at Advanced Information Networking and Applications, 2008. AINA 2008. 22nd International Conference, 2008.

[40] D. B. F. a. D. R. Cheriton, "DoS and Authentication in Wireless Public Access Networks," *In Proc. of the ACM Workshop on Wireless Security (WiSe'02)*, pp. 47-56, 2002.

[41] S. D. J. Kong, E. Tsai, and M. Gerla, "ESCORT: A Decentralized and Localized Access Control System for Mobile Wireless Access to Secured Domains," *In Proc. of the ACM Workshop on Wireless security (WiSe'03)*, pp. 61-68, 2003.

[42] A. K. Agarwal and W. Wang, "On the Impact of Quality of Protection in Wireless Local Area Networks with IP Mobility," *ACM Mobile Networks and Applications (ACM MONET)*, vol. 12, pp. 93-110, 2007.

[43] A. K. Agarwal, W. Wang, R. Gupta, and M.-Y. Chow, "LAP: Link-Aware Protection for Improving Performance of Loss and Delay Sensitive Applications in Wireless LANs," *In Proc. of IEEE Milcom'07*, pp. 420-425, 2007.

[44] A. S. M. Y. Matsunaga, T. Suzuki, and R. H. Katz, "Secure Authentication System for Public WLAN Roaming," *In Proc. of the 1st ACM International Workshop on Wireless Mobile Applications and Services on WLAN Hotspots*, pp. 113-121, 2003.

[45] I. Martinovic, F. A. Zdarsky, A. Bachorek, C. Jung, and J. B. Schmitt, "Phishing in the Wireless: Implementation and Analysis," presented at Proceedings of the 22nd IFIP International Information Security Conference (SEC 2007), Johannesburg, South Africa, 2007.

[46] J. Cheng, W. Haining, and G. S. Kang, "Hop-count filtering: an effective defense against spoofed DDoS traffic," in *Proceedings of the 10th ACM conference on Computer and communications security*. Washington D.C., USA: ACM, 2003.

[47] Rosenberg, "Request Header Integrity in SIP and HTTP Digest using Predictive Nonces," Internet Engineering Task Force, Internet Draft, 2001.

[48] I. Lee and R. Hunt, "A novel design of a VoIP firewall proxy to mitigate SIP-based flooding attacks," *Int. J. Internet Protocol Technology*, vol. 3, pp. 128-135, 2008.

[49] R. Laboratories, "RSA Cryptography Standard, http://ftp.rsasecurity.com/pub/pkcs/pkcs-1/pkcs-1v2-1.pdf," 2002.

[50] A. Juels and J. Brainard, "Client Puzzles: A Cryptographic Countermeasure Against Connection Depletion Attacks," *In Proceedings of the 1999 ISOC Network and Distributed System Security Symposium*, pp. 151-165, 1999.

[51] T. Aura, P. Nikander, and J. Leiwo, "DOS-Resistant Authentication with Client Puzzles," *In Revised Papers from the 8th International Workshop on Security Protocols*, pp. 170-177, 2001.

[52] M. C. Lee and F. Chun-Kan, "A public-key based authentication and key establishment protocol coupled with a client puzzle," *J. Am. Soc. Inf. Sci. Technol.*, vol. 54, pp. 810-823, 2003.

[53] L. Jussipekka, A. Tuomas, and N. Pekka, "Towards Network Denial of Service Resistant Protocols," in *Proceedings of the IFIP TC11 Fifteenth Annual Working Conference on Information Security for Global Information Infrastructures*: Kluwer, B.V., 2000.

[54] T. J. McNevin, J.-M. Park, and R. Marchany, "pTCP: A Client Puzzle Protocol For Defending Against Resource Exhaustion Denial of Service Attacks," *Technical Report TR-ECE-04-10, Dept. of Electrical and Computer Engineering, Virginia Tech*, 2007.

[55] W. XiaoFeng and K. R. Michael, "Defending Against Denial-of-Service Attacks with Puzzle Auctions," in *Proceedings of the 2003 IEEE Symposium on Security and Privacy*: IEEE Computer Society, 2003.

[56] A. Tuomas, N. Pekka, and L. Jussipekka, "DOS-Resistant Authentication with Client Puzzles," in *Revised Papers from the 8th International Workshop on Security Protocols*: Springer-Verlag, 2001.

[57] D. Drew and S. Adam, "Using client puzzles to protect TLS," in *Proceedings of the 10th conference on USENIX Security Symposium - Volume 10*. Washington, D.C.: USENIX Association, 2001.

[58] H. Dobbertin, "Cryptanalysis of MD4," *Fast Software Encryption: Third International Workshop*, pp. 53-69, 1996.

[59] R. Rivest, "RFC1321 - The MD5 Message-Digest Algorithm," *Network Working Group, MIT Laboratory for Computer Science*, 1992.

[60] "RFC 3174: US Secure Hash Algorithm 1 (SHA1)," 2001.

[61] Z. Rui, H. Goichiro, and I. Hideki, "A generic construction of useful client puzzles," in *Proceedings of the 4th International Symposium on Information, Computer, and Communications Security*. Sydney, Australia: ACM, 2009.

[62] D. McGrew, "Key Derivation Functions and their Uses," in *Internet-Draft, Network Working Group*, 2010.

[63] B. Kaliski, "RFC2898: Password-Based Cryptography Specification. Version 2.0," 2000.

[64] Clancy T, "Secure handover in enterprise WLANs: CAPWAP, HOKEY, and 802.11r," *IEEE Wireless Communications*, vol. 15, pp. 80-85, 2008.

[65] M. Kassab, J. M. Bonnin, and K. Guillouard, "Securing fast handover in WLANs: a ticket based proactive authentication scheme," presented at Globecom Workshops, 2007 IEEE 2007.

[66] Maccari L, Fantacci R, Pecorella T, and F. F, "Secure, fast handoff techniques for 802.1X based wireless network," presented at Proceedings of the IEEE International Conference, Istanbul, Turkey, 2006.

[67] V. Brik, A. Mishra, and S. Banerjee, "Eliminating handoff latencies in 802.11 WLANs using Multiple Radios: Applications, Experiences, and Evaluation," presented at Internet Measurement Conference, 2005.

[68] Kassab M, Belghith A, Bonnin J, and S. S, "Fast pre-authentication based on proactive key distribution for 802.11 infrastructure networks," presented at Proceedings of the 1st ACM WMuNeP, Montreal, Que., Canada, 2005.

[69] A. D. H. Duong, S. Gordon, "Proactive Context Transfer and Forced Handover in IEEE 802.11 Wireless LAN based Access Networks," *SIGMOBILE Mob. Comput. Commun.*, pp. 32-44, 2005.

[70] M. A, J. T, and A. D, "Minimizing re-authentication overheads in infrastructure IEEE 802.11 WLAN networks," presented at Wireless Communications and Networking Conference, New Orleans, 2005.

[71] M. S. A. Mishra, W. Arbaugh, "An Empirical Analysis of the IEEE 802.11 MAC Layer Handoff Process," *SIGCOMM Computer Communications*, vol. 33, pp. 93-102, 2003.

[72] W.A. Arbaugh, N. Shankar, J. Wang, and K. Zhang, "Your 802.11 network has no clothes," *IEEE Wirel. Commun. Mag*, vol. 19, pp. 44-51, 2002.

[73] *General Characteristics of International Telephone Connections and International Telephone Circuits*, vol. ITU-TG.114: International Telecommunication Union, 1988.

[74] Balachandran A, Velker GM, Bahl P, and R. PV, "Characterizing user behaviour and network performance in a public wireless LAN," presented at Proceedings of the ACM SIGMETRIC, California, U.S.A, 2002.

[75] M. S. A. Mishra, W. Arbaugh, "Context Caching using Neighbor Graphs for Fast Handoffs in a Wireless Network," presented at Proceedings of the IEEE INFOCOM Conference, Hong Kong, 2004.

[76] Pack S, Jung H, Kwon T, and C. Y, "SNC: a selective neighbor caching scheme for fast handoff in IEEE 802.11 wireless networks," *ACM Mobile Computing and Communications*, vol. 9, pp. 39-49, 2005.

[77] M. Kassab, J. M. Bonnin, and A. Belghith, "Fast and Secure Handover in WLANs: An Evaluation of the Signaling Overhead," presented at Consumer Communications and Networking Conference. CCNC 2008. 5th IEEE 2008.

[78] "IEEE 802.11f: IEEE Trial-Use Recommended Practice for Multi-Vendor Access Point Interoperability via an Inter-Access Point Protocol Across Distribution Systems Supporting IEEE 802.11 Operation," *IEEE*, 2003.

[79] "IEEE Std 802.11r - IEEE Standard for Information Technology-Telecommunications and Information Exchange Between Systems-Local and Metropolitan Area Networks-Specific Requirements Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications Amendment 2: Fast Basic Service Set (BSS)," *IEEE Std 802.11r-2008 (Amendment to IEEE Std 802.11-2007 as amended by IEEE Std 802.11k-2008)*, 2008.

[80] B. C. Bangolae S., Qi E., "Performance Study of Fast BSS Transition using IEEE 802.11r," presented at International Wireless Communication and Mobile Computing Conference (IWCMC' 06), 2006.

[81] Clancy T, Nakhjiri M, Narayanan V, and Dondeti L, "RFC 5169: Handover Key Management and Re-Authentication Problem Statement," 2008.

[82] Velayos H. and G. Karlsson, "Techniques to Reduce IEEE 802.11b MAC Layer Handover Time," in *Laboratory for Communication Networks, KTH, Royal Institute of Technology*. Stockholm, Sweden, 2003.

[83] M. Shin, A. Mishra, and W. Arbaugh, "Improving the Latency of 802.11 Hand-offs using Neighbor Graphs," *Proc. ACM MobiSys*, 2004.

[84] I. Ramani and S. Savage, "SyncScan: Practical Fast Handoff for 802.11 Infrastructure Networks," presented at Proceedings of the IEEE INFOCOM Conference, 2005.

[85] Haitao Wu, Kun Tan, Yongguang Zhang, and Q. Zhang, "Proactive Scan: Fast Handoff with Smart Triggers for 802.11 Wireless LAN," presented at IEEE INFOCOM proceedings, 2007.

[86] Marc Emmelmann, Sven Wiethoelter, and H.-T. Lim, "Opportunistic Scanning: Interruption-Free Network Topology Discovery for Wireless Mesh Networks,"

International Symposium on a World of Wireless, Mobile and Multimedia Networks 2009.

[87] P. Yogesh Ashok and A. Varsha, "Improving the IEEE 802.11 MAC layer handoff latency to support multimedia traffic," in *Proceedings of the 2009 IEEE conference on Wireless Communications \& Networking Conference*. Budapest, Hungary: IEEE Press, 2009.

[88] Sebastian Speicher and Christian Bunnig, "Fast MAC-Layer Scanning in IEEE 802.11 Fixed Relay Radio Access Networks," presented at International Conference on Mobile Communications and Learning Technologies, 2006.

[89] G. P. Pollini, "Trends in Handover Design," *IEEE Communications Magazine*, 1996.

[90] H. Ahmed and H. Hassanein, "A performance study of roaming in wireless local area networks based on IEEE 802.11r," *Communications, 2008 24th Biennial Symposium* pp. 253-257, 2008.

Glossary

AAA	Accounting, Authorisation, Authentication
AP	WLAN Access Point
APN	Access Point Nonce
AS	Authentication Server
BSS	Basic Service Set
BSSID	Basic Service Set Identity
CS	Channel Switching
dBm	Decibel-Mill Watt
DS	Distribution System
EAPoL	Extensible Authentication Protocol over Local Area Network
ESS	Extended Service Set
ESSID	Extended Service Set Identity
FATP	Fast AP Transition Protocol
IAPP	Inter Access Point Protocol (IEEE 802.11f)
IAT	Inter-Arrival Time
IEEE	Institute of Electrical and Electronic Engineers
IETF	Internet Engineering Task Force
IP	Internet Protocol
KH	Key Holder
LAN	Local Area Network
LM-SS	Location Management based Selective Scanning
MAC	Medium Access Control

NG	Neighbour Graph
NIC	Network interface card
PKI	Public Key Infrastructure
PMK	Pairwise Master Key
PMKID	Pairwise Master Key Identifier
PMKSA	Pairwise Master Key Security Association
PS	Power Save
QoS	Quality of Service
RADIUS	Remote Authentication Dial In User Service
RSN	Robust Security Network
RSNA	Robust Security Network Association
RSNIE	Robust Security Network Information Element
RSS	Received signal strength
RSSI	Received signal strength indicator
SIP	Session Initiation Protocol
SNR	Signal-to-noise ratio
SSID	Service Set Identifier
STA	WLAN Client Station
TLS	Transport Layer Security
TPC	Transmission Power Control
UDP	User Datagram Protocol
VoIP	Voice over Internet Protocol
WEP	Wired Equivalent Privacy
Wi-Fi	Wireless Fidelity
WLAN	Wireless Local Area Network

WNIC	Wireless Network Interface Card
WPA	Wi-Fi Protected Access